

BLOCK 1**Information Technology for Managers 5**

BLOCK 2**Information Systems 63**

BLOCK 3**System Analysis and Computer
Languages 143**

BLOCK 4**Support Systems for Management
Decisions 215**

EXPERTS COMMITTEE

Prof. K. Ravi Sankar,
Director, SOMS,
IGNOU, New Delhi

*Prof. Shubhamoy Dey
Department of Information System,
Indian Institute of Management, Indore

*Dr. Veena Bansal
Department of Industrial and
Management Engineering (IME),
Indian Institute of Technology,
Kanpur

*Sri Sanjay Jasola
Computer Division
IGNOU, New Delhi

*Sri Pankaj Khare
Commonwealth Educational
Media Center for Asia,
Commonwealth of Learning, New Delhi

Prof. Abid Haleem
Faculty of Engineering and Technology,
Jamia Millia Islamia,
New Delhi

Prof. H D Sharma
Former Prof & Head,
Pant Nagar Engineering College,
Pant Nagar

Prof. A. P. Verma
National Institute of Technology
Patna

Prof. Gokulananda Patel
Birla Institute of Management Technology
Greater Noida

Prof. B. Sudheer
Dept of Management studies,
Sri Venkateswara University,
Tirupati

*Dr. Sanjay Gupta,
Social & Rural Research Institute,
IMRB International, New Delhi

*Dr. M. Shanmugham
IGNOU Sub-Regional Center,
Madurai

*Sri Samir Srivastava
Computer Science & Engineering
Department,
Kamla Nehru Institute of Technology,
Sultanpur

*Sri Sanjay Mathur
Department of Electronics &
Communication Engineering,
College of Technology, Pantnagar

*Sri Ram Misra,
SIDBI Innovation & Incubation Center,
Indian Institute of Technology,
Kanpur

Prof. Kuldeep Singh Sangwan
Mechanical Engineering Department,
Birla Institute of Technology and Science,
Pilani

Dr VSP Srivastav
Head (Retd), Computer Division,
IGNOU, New Delhi

Prof Ajay
Department of Industrial & Production
Engineering,
G. B. Pant University of Agriculture &
Technology, Pantnagar

Prof. Raj K Jain
Professor (Retd), Vikram University,
Ujjain

Course Coordinator and Editor

Prof. Anurag Saxena,
Professor in Management,
SOMS, IGNOU
New Delhi

Note: A large portion of this course is adapted from the earlier MS-07 course and the persons marked with (*) are the original contributors of MS-7 Study Material.

PRODUCTION TEAM

Mr. Tilak Raj
Assistant Registrar
MPDD, IGNOU, New Delhi

January, 2022

© Indira Gandhi National Open University, 2022

ISBN : _____

All right reserved. No part of this work may be reproduced in any form, by mimeograph or any other means, without permission in writing from the Indira Gandhi National Open University.

Further information on Indira Gandhi National Open University courses may be obtained from the University's office at Maidan Garhi, New Delhi-110068 or visit University's Website <http://www.ignou.ac.in>

Printed and Published on behalf of the Indira Gandhi National Open University, New Delhi by Registrar, MPDD, IGNOU, New Delhi.

Printed at : M/s Educational Stores, S-5 Bulandshahar Road Industrial Area, Site-1, Ghaziabad (UP)-201009



ignou
THE PEOPLE'S
UNIVERSITY

Block

1

INFORMATION TECHNOLOGY FOR MANAGERS

UNIT 1

Information Technology: An Overview	7
--	----------

UNIT 2

Computer Systems and Smart Devices	18
---	-----------

UNIT 3

Computer Software	33
--------------------------	-----------

UNIT 4

Networking Technologies	48
--------------------------------	-----------

BLOCK 1: INFORMATION TECHNOLOGY FOR MANAGERS

Unit 1: Information Technology: An overview begins with a note about IT, its use world wide and its advancement. It discusses about various types of Information systems. It entails a discussion on the business perspective of information technology. It talks about the Internet as network of networks around the globe and its use in business applications and social media. Use of IT in decision-making systems has also been discussed in this unit.

Unit 2: Computer Systems and Smart Devices: enables you to be familiar with the components of a computer system. It discusses about CPU, Main Memory, Secondary memory, Input and Output devices, Networking and networking devices, Smart Devices, and plug & play devices. You will also comprehend the concept of networking.

Unit 3: Computer Software describes the types of operating systems and their functions. You will learn about system software (operating system, language translations and utility programs) and application software (developed using programming languages). The philosophy behind open-source software is also explained there.

Unit 4: Networking Technologies facilitates you to be acquainted with the data communication networks, the OSI model, and the communication protocol. You will be able to differentiate and identify among LAN, MAN, SWAN and WAN, Internet and Intranet. You will be able to understand the benefits to an organization due to networks. Client/server model and Cloud model of computing are also discussed in this unit.

THE PEOPLE'S
UNIVERSITY

UNIT 1: INFORMATION TECHNOLOGY: AN OVERVIEW

Structure

- 1.1 Introduction
- 1.2 Objectives
- 1.3 Definitions of IT
- 1.4 Types of IT systems
- 1.5 IT and its business perspective
- 1.6 Internet and its business applications
- 1.7 Computer Aided Decision Making
- 1.8 Summary
- 1.9 Unit End Exercises
- 1.10 References and Suggested Further Readings

1.1 INTRODUCTION

Information Technology (IT) has become a strategic necessity. Unless we believe in IT and act on it, there is every chance of becoming a footnote in the annals of History. IT has become a vital component of successful businesses and organizations. Managers are expected to identify opportunities to implement Information System (IS) to improve the business processes. Managers are also required to lead IS projects in IT.

Information Technology has become a major facilitator of business activities. It is also a catalyst of fundamental changes in the structure, operations and management of organizations. IT can be used to:

1. Perform high-speed, high-volume, numeric computations.
2. Provide fast, accurate and inexpensive communication within and between organizations.
3. Store huge amounts of data in an easy-to-access yet small space and allow quick and easy access.
4. Automate business processes and manually done tasks.
5. Analyse big volume of data available through internet for decision making.

IT has been used for improving productivity, reducing cost, enhancing decision making process, enhancing customer relationships, and developing new strategic applications. The business paradigm has completely shifted from being sellers market to buyers market. Customers have become the focal point of any business. The business environment is no more as stable as it used to be and has become much more competitive. It became mandatory on the part of the organizations to make full use of IT to survive. IT has become one of the standard components of an organization.

The individuals are supported by IT to fulfill their roles. The management and the business processes have become IT oriented. Organization structure and strategy are also supported by IT. Whenever an external or internal pressure is felt by an organization, IT helps the organization to plan critical response activities. The changed government policy may create a pressure on an organization. But such a pressure is distributed over a long period as government gives enough time to organizations to respond to changed policies. But if there is a change in the behavior of consumers, the organization should be able to identify the change. Moreover, they should be able to come up with a solution fast. IT helps an organization to anticipate and stay ahead of problems. The organization can take proactive measures rather than firefighting measures.

An organization can use an IT supported strategic system to increase their market share.

IT can help an organization negotiate better with their suppliers. Maintaining and improving quality of products and processes in an organization needs regular support, vigilance, and innovation. IT has been used extensively for productivity improvement, reducing inventory and maintaining quality. Management Information System (MIS) and decision support system (DSS) have been used to help management in decision-making process.

Business process reengineering (BPR) has become the need to the day for every business; BPR involves changing business processes in an innovative way. IT plays a major role in BPR. Internet and Intranet help an organization in changing its business processes to reduce cycle time and time to market a product. IT makes information available to employees with different level of access. As a result, employees can be given more independence to make decisions. The ERP, a strategic tool, heavily uses IT to integrate business processes of an organization.

1.2 OBJECTIVES

After reading this unit, you should be able to:

- Define information technology and state the advancement of IT;
- Identify the different types of Information systems;
- Contextualize a business perspective of IT;
- Describe the significance of Internet and its business applications;
- Describe and understand the wide coverage of Electronic Commerce and
- Explain the role of Computer Aided Decision System in business and other sectors and services.

1.3 DEFINITIONS OF IT

Let us understand what information is. *Information is the finished product for which data is the raw material.* The dictionary defines information as *processed data, which is used to trigger certain actions or gain understanding of what the data implies.*

Information has also been defined as *data that have been put into a meaningful and useful context and communicated to a recipient who uses it to make decisions. Information involves the communication and reception of intelligence or knowledge. It appraises and notifies; surprises and stimulates, reduces uncertainty, reveals additional alternatives or helps eliminate irrelevant or poor ones, and influences individuals and stimulates them to action.* The information must be received by the recipient within the required time frame and the information must be free from errors.

The technology plays an important role in delivering timely and error free information to its recipients. Technology includes hardware, software, databases, and communication system. Hardware is a set of devices such as processor, monitors, keyboard, and printer that accept data, process them, and display them. Software is a set of programs that enable the hardware to process data. Database is also an integral part of IT system, which is a collection of related files, tables, relation etc. that stores data and the association among them. Network connects computing resources of an organization and facilitates sharing of hardware and software. The organization processes and people are integral part of an IT System.

Now we know what information is and what technology is. With this, we are ready for a definition of IT.

Information Technology means the collection, storage, processing, dissemination, and use of Information. It is not confined to hardware and software but acknowledges the importance of person and the goals he/she sets for his/her technology, the values employed in making these choices, the assessment criteria used to decide whether he/she is controlling the technology and is being enriched by it.

Initially, the Information Technology Serving Society, USA gave the above definition, way back in 1979. It has been further generalized with wide coverage.

The above definition clearly states that IT is an important tool, which must be used properly. At one time, 60% people used to work in agriculture. Nowadays, in a developed country such as US, more than 91% households have computers and Internet connection. In India, this figure is about 54%. Thus only 540 people out of every 1000 have access to computers. However, India is making steady progress because of IT use in almost every field.

As far as history of computing is concerned, people tried to invent a computing machine as early as 800 BC. Abacus is one of the oldest computing devices that are still in use. Every child in China learns to use abacus in school. Our ancestors had built various other mechanical machines using gears. All these efforts laid foundation for better computing machines. The diode tubes, transistors, integrated circuits (ICs) and now very large integrated circuits (VLSI) are the electro-mechanical devices that have been invented over a period of time. The technology has improved since World War II many folds. The improved technology has been the key factor in making better computing devices. Along with improvement in computing technology, the communication technology has also improved in parallel. The software also improved and became user friendly. The spreadsheets, the

word processing packages, database packages, simulation and automation software packages, and decision support systems made IT popular among managers as well. The terms Information Systems (IS) and Information Technology (IT) are used synonymously. IS has evolved considerably since 1960.

Activity A

Give examples from day-to-day activities in your organization or any organization of your choice to support the statement, “Managers need only information and not data”.

.....

.....

.....

.....

1.4 TYPES OF INFORMATION SYSTEM

There are various types of information systems as listed below.

Transaction Processing Systems (TPS)

A TPS is used primarily for record keeping which is required in any organization to conduct the business. Examples of TPS are sales order entry, payroll, and shipping records etc. TPS is used for periodic report generation in a scheduled manner. TPS is also used for producing reports on demand as well as exception reports.

Decision Support System (DSS)

DSS serves the management of an organization. A decision support system has sophisticated data analysis tools, which support and assist all aspects of problem specific decision making. DSS may use data from external sources such as current stock prices to enhance decision-making. DSS is used when the problem is complex and the information needed to make the best decision is difficult to obtain and use. DSS is developed with the help of decision makers. DSS helps in decision-making process and does not make any decision.

Executive Information System (EIS)

An Executive Information System is also called Executive Support System. Senior managers of an organization use the EIS. Therefore, it must be easy to use so that executives can use it without any assistance. EIS can do trend analysis, exception reporting and have drill down capabilities. The results are usually presented in a graphical form tailored to executive’s information needs. EIS has on-line analysis tools and they access a broad range of internal and external data.

Management Information Systems (MIS)

MIS provides the management routine summary of basic operations of the organization. The basic operations are recorded by the TPS of the organization and MIS consolidates the data on sales, production etc. MIS provides routine information to managers and decision makers. The

main objective behind installing an MIS in the organization is to increase operational efficiency. MIS may support marketing, production, finance etc.

Work Flow system

A workflow system is a rule based management system that directs, coordinates, and monitors execution of an interrelated set of tasks arranged to form a business process. A workflow system is also known as a document image management system. For example, a workflow system is used by banks for loan sanction process. An applicant fills out an electronic application form for a bank loan at a bank's web site. The application is then upload to the bank loan officer's site. The loan officer conducts an interview and fills in his feedback and passes the application form to the credit check unit. The credit unit checks the credit limit and fills in the details. The application is now complete and a final decision is made.

There are three types of workflow software. Administrative workflow systems focus on the tracking of expense reports, travel requests, messages. An Ad-hoc workflow system deals with the shaping of product, sales proposals, and strategic plans. Production workflow systems are concerned with mortgage loans and insurance claims. A workflow system may be Internet based and may be combined with e-mail. A workflow system may be based on client/sever architecture that may use a database/file server.

Enterprise Resource Planning (ERP)

ERP system is a set of integrated programs capable of managing a company's vital business operations for an entire multi-site, global organization. This facilitates organizations to share and integrate data for consistent flow of information over the organization's network and also using the internet as the organization may decide.

Expert Systems

The system has the ability to make suggestions and act like an expert in a particular field.

An expert system has an extensive knowledge base. Nowadays such systems are in use for automating business processes to great extent. It is envisaged that expert systems will be in high demand further in future.

1.5 BUSINESS PERSPECTIVE OF IT

The impact of IT on a business has been tremendous. One of the advantages of IT systems for a business is the cost-performance ratio, which is better in case of computers. The labor cost increases every year but the cost of computer does not increase. A better and more powerful computer can be bought for the same price after a year. It is better to use computers for routine jobs as far as possible. The IT has been used in every business and for every function of a business. Some of the applications are as follows.

- **Finance and accounting** - IT has been used for forecasting revenues, determining the best sources and uses of funds and managing cash and other financial resources. IT has also been used to analyze investments and perform audits.

- **Sales and marketing** - IT has been used to develop new services, which may not exist without IT. IT has helped management of various organizations to determine the best location for production and distribution facilities. The operational data has been analyzed using IT to determine the best advertising and sales approaches. The product prices have been set using IT to get the highest total revenues. In other words, IT has been used for product analysis and price analysis.
- **Manufacturing** - IT has been extensively used for processing customer orders, controlling inventory levels, developing production schedules and for monitoring product quality. A whole new discipline- Computer Aided Design and Computer Aided Manufacturing has evolved due to application of IT to design and manufacturing. The manufacturing is not what is used to be due to the use of computers, Computer Integrated Manufacturing (CIM) dominates the manufacturing sector. Use of robots is increasing not only in manufacturing plants but also in other sectors such as healthcare, laboratories, defense sector etc.
- **Human Resource Management** - Companies are using IT systems for screening applicants and conducting various tests.
- **Project Management** - A range of software packages are available in the market for managing projects. These software products let the management set the schedules, milestones, facilitate communication among group members, and monitor the project progress. These products help in document and report preparation.
- **Data Analysis**- Investment firms heavily use information systems to analyze stocks, bonds and options to provide better service to their clients. With the internet connectivity this application is high in demand even by other firms dealing with large data.

Activity B

Give examples to highlight the applications of IT in

- Public Relations
- Market research
- Purchasing
- Healthcare
- Education

.....

.....

.....

.....

1.6 INTERNET AND ITS BUSINESS APPLICATION

The network technology had been developed during 1970s. The network hardware and software improved as a result of research and investment. The network became a reality and every organization laid network cables

and connected their computers and other resources to the network. Such a network is called Intranet. An Intranet is restricted to an organization. When a network of networks is formed, it is called Internet. In other words, Internet is a global network of computer networks. The Internet connects computing resources of various organizations such as academic institutes, business organizations and government organization. All networks, which are part of the Internet, follow a protocol called TCP/IP protocol for communication. Internet has affected business and people both. The Internet provides fast and inexpensive communication channels. The Internet is used for transferring data files, e-mail messages and for sharing documents and images. Internet is also used for chatting, social media, and networking. People, sitting miles apart, work on the same project making use of Internet to exchange ideas in real time. The Internet is also used for education and entertainment.

The Internet has changed the way business is done. A new business paradigm termed electronic commerce has come into existence. We will discuss e-commerce in some detail little later. Consumers are able to shop for goods and services from all over the world in the comfort of their homes. The individuals are able to shop, bank, work, and entertain themselves without leaving their homes. The payments are also made through Internet. The organizations, which provide these services, also use Internet. These organizations use Internet to conduct electronic meetings, train employees in many different locations simultaneously. The manufactures and corporate houses directly deal with the producers without going through the retailers. A large percentage of people are still not part of Internet. However, Internet users are increasing rapidly.

Internet has facilitated development of workflow systems across networks. As we have already discussed, workflow systems are business process automation tools that place system controls in the hands of user departments. They are very flexible and can be designed to automate almost any information-processing task. The primary purpose of workflow systems is to provide users with tracking, routing, document imaging, and other capabilities designed to improve business processes.

Teleconferencing, video conferencing and screen sharing are some of the other Internet applications. Tele-video conferencing save travel time and travel cost. The ideas can be shared quickly and the information flow is much faster. As a result, the product development time and contract negotiation time have reduced. The customer service has improved due to faster and accurate information availability. Many organizations have been able to draw competitive advantage by using Internet.

The Internet connects hundreds of thousands of different networks around the world. Today, almost each and every person working in academic institutions, business organizations, government organizations and many more like healthcare, R&D and so on use the Internet. Even the agriculture sector uses this tool for connecting and disseminating information. Uses of Internet vary from being a communication medium to providing a means for collaboration work.

The Internet is extremely robust. A system or a network can be added / deleted to the Internet without adversely affecting others on the Internet. The

communication media used by Internet can be a telephone line, an Ethernet/OFC cable, or a microwave link. The Internet can carry data, images, text, voice, videos or transactions without any discrimination.

The Internet is being used by various organizations for collaborative work. These organizations use the Internet to send electronic mail, message, to transmit documents and data. The organizations are organizing and participating in electronic conferences. Virtual conferences and on line classes/meetings/purchases/deliveries have proved to be the only means for survival of industries in the pandemic (Covid-19) worldwide.

Another popular use of Internet is to access information that others make available in public domain. There are special software package known as search engines, which help users locate the desired document. Large databases, electronic brochures, book details, and manuals are present on the Internet. Organizations advertise their products and services on the Internet.

Internet is also being used to participate in online discussions. These discussions are done in real time. An applicant can be interviewed online from a distant location. The candidate does not have to travel to appear for the interview. Apart from these obvious uses, the Internet is inspiring new business models, called digital firms. In digital firms, hierarchy and organizational levels are less compared to a traditional firm. The employees at lower-levels have access to more data and have more decision-making authority. They could be at location A and working for a digital firm at location B. These firms are much leaner and efficient than traditional firms.

Internet has immensely helped the field workers. In a traditional environment, a field worker had to go to his office to take orders for the day and report at the end of the day again to provide details of his/her accomplishments. An employee of a digital firm carries a laptop and enters the data into the laptop, which is transmitted to the organization. An employee can attend meetings without actually traveling to his/her office.

The existing business processes have been thoroughly redesigned to take advantage of IT. A workflow system has reduced the time drastically. There is no paper movement and everybody can work on an application in parallel.

A company can use global data along with its operational data to respond to changes in the market place. A company works with much less inventory as it may easily co-ordinate production activities in the light of the orders. The products can be customized. The customization done to cater to the needs a very small segment is referred to as micro marketing.

Internet has also been used for electronic commerce. In Electronic Commerce (EC), business transactions take place via telecommunication networks. The business transactions may take place between an organization and consumers or between two businesses or organizations. The major benefits to the company include reduced cost, reduced cycle time, and improved customer service. Following is a list of benefits to an organization using electronic commerce.

- EC decreases the cost of creating, processing, distributing, storing, and reliving information.

- EC allows reduced inventories and overheads.
- EC reduces the time between the payments and receipts of goods and services.

EC enables an organization to operate in the areas much beyond their physical location. It is the process of buying and selling goods and services electronically using Internet. The activities such as marketing, advertising, customer support, delivery, and payments are also done electronically. An e-business is supported by Internet within the organization.

One must keep in mind that e-commerce is a whole new business paradigm which needs fresh thinking. Just creating a web site doesn't guarantee success. Many people started e-business and failed. A sound e-business model is required. Some of the Internet business models are as follow:

Virtual storefront is an e-business where physical goods or services are sold online instead of an actual retail outlet. Some of the examples are www.amazon.com, www.bigbasket.com, grofers.com, www.swiggy.com which sells so many items online. A customer can select an item from their web site and place an electronic order. Payment can be made online through credit/debit card, net-banking and through other mode of online payments like UPI (Unified Payments Interface), QR (Quick Response) scanning, and Apps like PayTm, GooglePay etc. and then the company delivers goods to the delivery location using traditional means. There are e-businesses whose prime business is to sell items where prices are not fixed. The prices are decided by the customers through auction. There are e-businesses who are content provider and manager. The companies where core competence is not IT hire an IT company to create web sites and databases for them. Almost all companies provide portals for making business. A portal is a web site, as given above, provides an initial entry point to the web site of the company or other respective services.

The e-commerce can also be classified based on the parties involved in the business. A business-to-consumer e-business involves an electronic retail home and customers. Such a business is also called B2C business. www.amazon.com is an example of B2C business. Business-to-business (B2B) involves transaction among businesses. The main advantage is that a business house can buy raw material and supplies from another business party without going through the retailer. In a Consumer-to-Consumer business (C2C), the e-business is a facilitator only. People are able to sell to other people using a C2C facility. Governments are also providing services to their citizens under G2C (Government-to-Citizens) mode of facilities.

1.7 COMPUTER AIDED DECISION MAKING

A decision system supports and assists all aspects of problem specific decision-making.

A decision support system (DSS), also called Computer Aided Decision System (CADS), is used when the problem is complex and the information needed to make the best decision is difficult to obtain and use.

In an organization, a manager has certain goals, which he tries to achieve through the use of resources. The resources such as people, money, material,

and time is always limited. One of the roles that a manager plays among many others is of decision-making. The manager would like to examine various alternatives, which may not be possible in a manual system due to time constraint or due to sheer volume of data. Sometimes, a statistical analysis of fluctuating data is required which may be done only with the help of a decision support system. The data may be distributed over various branches of the organization and without the help of Information Technology; it may not be possible to examine the data.

An information system that helps a manager in making a decision is called a computer-aided decision support system (CADSS) or simply decision support system (DSS). A DSS is designed with the help of decision makers. A DSS aids in decision-making process and does not make any decision. A DSS may be used for doing a thorough risk analysis of a project. Such a DSS performs what - if analysis. A DSS consists of a model of the real world, collection of facts- database, and a user interface.

An analyst talks to the manager and tries to establish his requirement. Analyst may need to interact with users many times before the requirements are completely understood. The analyst may also study the existing system if any. The analyst must have domain knowledge. The reality is examined, the problem is identified, and it is defined. The problem may be too complex to be analyzed. In such a situation, a simplified version of the problem is defined. The simplified problem is used to create a model of the actual problem. There are many ways of creating a model for a problem. The model may be a scale model, i.e. model of a building or a bridge. It may be a mathematical model such as equations describing the trajectory of a missile. The model may be an analog model such as a blue print for a building or a map to show a particular region. There are many other types of models, which are not discussed here. A model has a set of uncontrollable variables that are not under the control of the manager. There is a set of decision variables, which are under the control of the manager, and these variables describe alternative course of action. As the values assigned to decision variables are changed, the results also change. The result variables are dependent on decision variables.

The analyst then documents the requirements comprehensively. For formal and pictorial representations diagrams such as Data Flow Diagram, Flowcharts, ER diagram, Class and Object diagram, structured English are used. Such a document is called Software Requirements Specification (SRS).

A system is then built according to the requirements. There are many software development models such as waterfall model, prototype, and incremental models. For a decision support system, usually a prototype is built which is shown to the manager for approval. The complete system is then built taking the feedback from the manager into consideration.

The decision variables are changed and impact on result variables is analyzed. A DSS provides support for decision makers by bringing together human judgment and computerized information. A DSS may provide support throughout large and complex organization or it may support an executive. A DSS may help an executive to perform trend analysis. A DSS

may be designed to generate exception report so that the manager does not miss any exceptional condition.

1.8 SUMMARY

In this unit, we have learnt about the Information Systems. The terms Information Systems and Information Technology are used synonymously. We learnt a few definitions of IT. We also learnt the main components of an IT system. We also learnt various types of IT systems that are used in a business. The business applications of IT were discussed in some detail. Internet and E-commerce which is one of the applications of Internet were discussed. Various business models of e-business were also discussed. IT has also been used extensively in decision-making systems. These systems do not make any decisions by themselves but they only help the decision maker in analyzing the data.

1.9 UNIT END EXERCISES

1. Define an information system and list its major components.
2. Describe how information system can support a business.
3. Define the Internet, Intranet, and e-commerce.
4. Briefly describe various types of Information Systems.
5. What is the purpose of a TPS? How does it compliment MIS in an organization?
6. List the major advantages of the Internet. Briefly describe impact of Internet on business.
7. Write a note on e-commerce. What are the advantages of e-commerce over conventional business?
8. What is the purpose of a computer aided decision support system? What are the major components of such a system? How does it help the decision making process?

1.10 REFERENCES AND SUGGESTED FURTHER READING

Lauden, K. C. and Laudon, J. P., (2002) *Management Information Systems- Managing the digital firm*, Pearson Education.

Turban, E., McLean, E. and Wetherbee, J., (1999), *Information Technology for Management*, John Wiley, and Sons Inc. (Asia).

<https://www.statista.com/topics/2237/internet-usage-in-the-united-states/#dossierKeyfigures> , site accessed on November 14, 2021.

UNIT 2: COMPUTER SYSTEMS AND SMART DEVICES

Structure

- 2.1 Introduction
- 2.2 Objectives
- 2.3 Components of a computer
- 2.4 Central Processing Unit (CPU)
- 2.5 Storage and storage devices
- 2.6 I/O devices
- 2.7 Networking and networking devices
- 2.8 Plug and play devices
- 2.9 Smart Devices
- 2.10 Communication technology
- 2.11 Types of computer
- 2.12 Summary
- 2.13 Unit End exercises
- 2.14 References and Suggested Further Readings

2.1 INTRODUCTION

For an information system, hardware is defined as any machinery that helps in input, processing, storage, and output activities. Similarly for a computer, the hardware is the collection of devices that perform the functions of input, processing, data storage, and output. In other words, all physical units of a computer system constitute computer hardware. The input device gets the data from the outside world and the data is stored in the memory. The central processing unit (CPU) processes this data and the various output devices provide the results. The components communicate with each other through system bus. Each hardware component plays an important role in computing. The arrangement of the component within the system even today is what was suggested in 1945 as stored program concept of computing by Von Neumann and is known as Von Neumann architecture. Smart devices use internet or the organizational network and work as information processors and information providers. A smart device is an electronic device, generally connected to other devices or networks via different wireless protocols such as Bluetooth, Wi-Fi etc., that can operate to some extent interactively and autonomously. They can be used in almost any industry, from smart manufacturing to healthcare, helping to improve efficiency and optimize operations.

2.2 OBJECTIVES

After reading this unit you should be able to:

- Identify the components of a computer system;

- Describe the elements of a CPU;
- Explain the concept of networking;
- Use plug and play devices; and
- Understand about smart devices.

2.3 COMPONENTS OF A COMPUTER

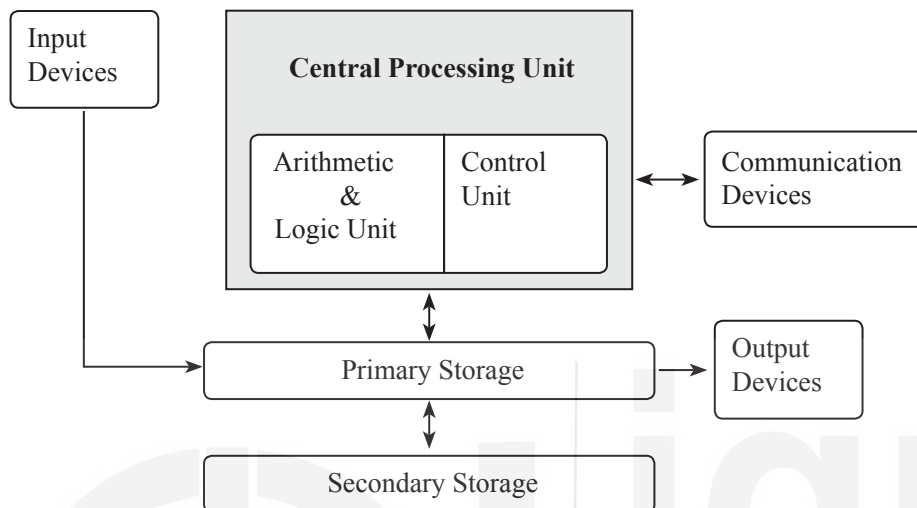


Fig 2.1: Components of a Computer

The computer memory is categorized into primary storage and secondary storage. Primary storage stores the data and programs during processing and secondary storage is used to store the data and programs for future use. Communication devices connect a computer to other computers and devices to facilitate data exchange with them.

Computers have been around since 1940s and have been evolving ever since. Computer hardware has mainly evolved through four stages.

The so-called first generation (1946-1956) computer was built by vacuum tubes and programming was done by plugging and unplugging chords. The second-generation (1957-1963) computers had transistors. The third generation (1964-1979) computers had Integrated Circuits (ICs). The fourth generation (1979 onwards) systems use Very Large-Scale Integrated (VLSI) circuits. Along with the hardware technology, the software also improved in terms of being user friendly and capability. Every system has software as its integral part, which makes the computer work. The most important software is an operating system (OS), which performs two functions:

- Extend the machine - OS presents to the user an extended machine that is easier to program than the underlying hardware.
- Manage resources such as processor(s), memories, timers, disks, mouse, network, printer etc. OS provides an orderly and controlled allocation of the resources among the various programs competing for them.

2.4 CENTRAL PROCESSING UNIT

In order to work, a computer needs some sort of “brain”. At the core of each computer, there is a device called central processing unit (CPU), which is

the brain of the computer. CPU reads the program from the main memory, and executes each step of the program, which may involve calculations and decision-making. The CPU is responsible for controlling all devices of the computer through its component – the Control Unit. It initiates a memory operation, which may involve reading data from an input device and storing it into memory or read data from the memory and display it on an output device.

The CPU mainly consists of three parts- Control Unit, Arithmetic and Logic Unit (ALU), and its storage also referred to as the CPU registers.

The ALU is the unit, which performs all mathematical calculations and logical operations. It performs addition, subtraction, multiplication, and division. It performs a logical operation by comparing two numbers. It can determine the smaller number, larger number or determine if the two numbers are equal. It can also determine whether a number is positive, negative or zero.

A program is a sequence of instructions. An instruction may be an arithmetic operation, a logical operation, an assignment, or a jump. The control unit sequentially accesses program instructions, decodes them, and directs ALU, Main Memory also referred to as Primary Storage, input devices, and output devices so that the program instructions can be carried out. Execution of one program instruction may require control unit to issue many directives. The ALU may also perform many operations to complete one program instruction. Each operation performed by ALU and control unit is referred to as machine instruction. Each program instruction may involve many machine instructions. A single machine instruction is completed in a machine cycle. The number of machine instructions completed in a second is called the speed of the CPU and it is measured in MIPS (Millions of Instructions Per Second) or BIPS (Billions of Instructions Per second). A CPU also contains a set of registers, which are specialized, small, high-speed memory for storing temporary results of ALU and for storing control information.

Computer Speed

The word length of a CPU is the number of bits it can process in a single cycle. A 64-bit machine can process 64 bits in a single cycle. A 64-bit machine is faster than a 32-bit machine. Today as fast as 256 bit processors are available which are much faster than what we used to have earlier.

The cycle is defined by the internal clock of the control unit. We have systems with as much as 5.1 GHz (Giga Hertz). Intel Core i7 & Core i9 based systems provide this speed. Similarly it may be available with other manufacturers like AMD also. In yester years, systems used to be with KHz (Kilo Hertz) and MHz (Mega Hertz).

The bus carries data, control signals, and address in a system. If the data bus width is same as the word length, then one word can be moved at a time. If the bus width is half of word length, two cycles are required for moving one word.

The instruction set also affects the speed. If the instruction set is simple, as in RISC (Reduced Instruction Set Computers), one or more instructions

may get executed in each cycle. There will be some long instructions, which take more than one cycle. It is possible that a given application extensively uses long instructions (such as scientific computing) and the machine will appear to be slow.

There is no direct relationship between clock frequency and the speed of a system. There are benchmark programs that are run to establish the speed of a computer system.

2.5 STORAGE AND STORAGE DEVICES

The large amount of data is stored on a computer using various types of storage media. The storage media are distinguished by their relative speed, capacity, and resilience to failure.

- 1) *Volatile Storage*: Information residing in such storage needs continuous power supply. The contents are lost if power supply is switched off. Examples of such storage devices are main memory and cache memory, which we will discuss in the next section. Access to volatile storage is very fast, both because of the technology used and because of the access method.
- 2) *Nonvolatile Storage*: The nonvolatile storage media do not require power supply to retain their contents. Examples of such storage media are disks and magnetic tapes. Disk is used for online storage, while tapes are used for archival storage. Disks and magnetic tapes are very reliable storage media. The current technology used for nonvolatile storage makes them much slower than volatile storage.

Main Memory

Main memory holds the programs and data required by the CPU for carrying out its operations. The primary storage is a semiconductor device that is built using integrated circuits. The data is stored in binary form by main memory. Numeric as well as non-numeric data can be represented in binary form. With two binary digits, we can represent 4 different characters. With three binary digits, we can represent 8 different characters. Computers internally use eight binary digits to represent characters and digits (A binary digit is referred to as bit and 8 bits are called a byte). 256 characters can be represented by a byte. The main memory consists of many thousands of bytes. The table given below lists commonly used names, abbreviations and the number of bytes for storage capacity.

Table 2.1: Commonly Used Names and Abbreviations for Storage Capacity

NAME	ABBREVIATION	NUMBER OF BYTES
Byte	B	1
Kilobyte	KB	1,024
Megabyte	MB	1,024 * 1,024 (about one million)
Gigabyte	GB	1,024 * 1,024 * 1,024
Terabyte	TB	1,024 * 1,024 * 1,024 * 1,024

Types of Main Memory

Memory can be of various types like Random Access Memory (RAM) and Read-Only Memory (ROM). The figure 2.2 given below gives a description of the memory types.

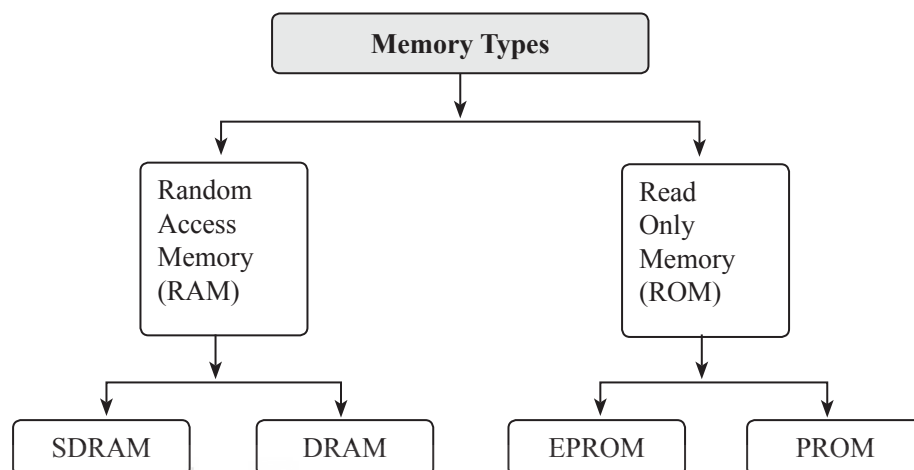


Figure 2.2: Types of Memory

RAM has become the synonym for main memory. Earlier, the word core memory was used for referring to main memory. RAM is the memory access method and core memory was the technology used in main memory. The core memory technology is not used any more. RAM is an acronym for Random Access Memory. As the name suggests, any location of the memory can be accessed randomly and the access time to is independent of the location. We will continue to use word RAM to refer to main memory. RAM is very fast; the access time is in nano seconds. RAM is volatile that is the contents of RAM are lost when the power supply to the RAM is discontinued.

DRAM (Dynamic Random Access Memory) has been the most common kind of RAM. The data is stored in the cell of transistors and capacitors and the data has to be refreshed every few milliseconds. SRAM (Static Random Access Memory) does not require periodical refresh. SRAM is faster than DRAM but is more expensive as compared to DRAM. SDRAM (Synchronous DRAM) is used these days. In fact the first version of SDRAM which is also known as DDR (Double Data Rate) RAM came in 1998. Further advancements in DDR have gone up to DDR5 (July 2020) which has capacity of double the capacity of DDR4 i.e. 2x21.3GB, the DDR4 being of 21.3GB.

A part of computer storage is ROM that cannot be erased or changed. ROM is an acronym for read only memory. ROM is nonvolatile i.e. its contents are not lost when power is switched off. ROM is required for storing the boot program that should not be lost or changed due to any failure. ROM also comes in many flavors such as PROM and EPROM. PROM (Programmable Read Only Memory) is used for storing some specialized application by the computer designers instead of chip designers. PROM can be written only once. EPROM (Erasable Programmable Read Only Memory) can be erased and reprogrammed many times.

Secondary Memory

Main Memory provides a small amount of volatile storage. But a computer needs to store large amount of data and instructions permanently. The secondary memory is non-volatile and has large capacity. The secondary memory is slow as compared to main memory. Various technologies are used for secondary memory, some provide random access, and others provide sequential access. In sequential access, data must be accessed in the order in which it is stored. For example, if we have stored data of students in the order of their roll numbers, then to retrieve data of student with roll number 20, we need to read (and possibly discard) the data of students with roll numbers 01 to 19 before we can access the desired data. If direct access is provided then any part of the data can be accessed directly, without the need to pass by other data in sequence. The sequential access storage devices are referred to as SASD (Sequential Access Storage Devices) and the direct access devices are called DASDs (Direct Access Storage Devices). The most common forms of secondary storage devices are magnetic tapes and magnetic disks.

Magnetic tape is one of the common sequential secondary storage mediums. The tape usually is a Mylar film coated with iron oxide. Portions of the tape are magnetized to represent bits. It is sequential access device, to access the n th block on tape; we must first read the preceding $n-1$ blocks. The main advantages of the magnetic disks are that they are inexpensive, long lasting and can store large volume of data (in Tera Bytes). Generally it is used for back up or archival storage of data.

Magnetic Disk is all made of magnetic material shaped as a thin circular disk and protected by a plastic or acrylic cover. A disk is single sided if it stores information on only one of its surfaces and double sided if both surfaces are used. To increase storage capacity, disks are assembled into a disk pack, which may include many surfaces. Information is stored on the disk surface in concentric circles of small width, each having a distinct diameter. Each circle is called a track. For disk packs, the tracks with the same diameter on the various surfaces are called a cylinder. The concept of cylinder is very important because data stored on the same cylinder can be retrieved much faster than if it were distributed among different cylinders.

Each concentric circle typically stores the same amount of data. The number of tracks on a disk ranges up to 1023. Each track is divided into sectors. A sector can store up to 4KB, available in laptop computers. The division of a track into equal sized blocks or pages is set by the operating system during disk formatting. There is a read/write head, which reads data from the disk and writes data to the disk. A disk is mounted on the disk drive, which has the motor that rotates it. A read/write head has a mechanical arm attached to it. All arms are connected to an actuator attached to another electrical motor, which moves the head in unison and positions it over the cylinder of tracks specified in a block address.

RAID (Redundant Array of Independent Disks) is an array of multiple inexpensive hard disks that provides fault tolerance through redundancy and improved access rates. It provides a method of accessing multiple individual disks, as they are one large disk. The data is spread over these multiple

disks, thereby reducing the risk of losing all data if one disk fails. Usually, RAID is used in a computer, which is a file server or a transaction server, where data accessibility is critical and fault tolerance is required. RAID is also used in desktop systems for CAD, multimedia editing, and playback systems where higher transfer rates are needed. RAID levels storage can be used for three distinct purposes, including:

Capacity: Capacity is a simple one that is especially involved in most every type of RAID setup used. For example, two hard drives can be linked together as a single drive to the OS effectively making a virtual drive that is twice the capacity.

Security: RAID levels can be used for data security by using some of the space on the drives to essentially clone the data that is written to both drives. Once again, with two drives you can make it so that the data is written to both drives.

Performance: Performance is another key reason for using a RAID setup on a personal computer. Also, RAID levels can be used for data security by using some of the space on the drives to essentially clone the data that is written to both drives. Once again, with two drives you can make it so that the data is written to both drives.

The RAID has following levels:

- 1) *RAID 0:* It is also known as “Disk Striping”. This is technically not a RAID level since it provides no fault tolerance. Data is written in blocks across multiple drives for fast access. The advantages of striping are the higher access rate, and full utilization of the array capacity. The disadvantage is there is no fault tolerance
- 2) *RAID 1:* This is also known as “Disk Mirroring”. Raid 1 provides redundancy by writing the data twice. If one disk fails, the other contains an exact duplicate of the data and the RAID can switch to using the mirror drive with no lapse in user accessibility. The disadvantages of mirroring are no improvement in data access speed, and higher cost, since twice the number of drives is required. However, it provides the best protection of data since the array management software will simply direct all application requests to the surviving disk members when a member of disk fails.
- 3) *RAID 2:* In this type of RAID set up disk striping occurs at the bit-level. RAID 2 is seldom deployed because cost to implement are usually high as a typical set up requires 10 disks and gives poor performance with some I/O operations.
- 4) *RAID 3:* RAID level 3 distribute data across multiple discs and stores additional information on one disc for error correction in recovery. It has high reliability and it is fast.
- 5) *RAID 4:* RAID 4 is a configuration in which disk striping happens at the byte level rather than bit level as in RAID 3.
- 6) *RAID 5:* It is the most popular configuration. In RAID 5, the additional data, which is used for error detection and correction, is also distributed over many discs. This arrangement has very high reliability and access speed is also good. However, the write operation

is slower compared to other RAID configurations.

- 7) **RAID 6:** RAID 6 is also used frequently in enterprises. It is identical to RAID 5 except that it is an even more robust because it uses one more parity block than RAID 5. You can have two disks non-functional (dead) and still have a system be operational.

There are further RAID levels possible but above are sufficient for your knowledge.

Optical Disks are the disks made of plastic, in which lasers that physically burn pits in the disk record data. Here each pit represents a binary digit 1 and each unpitted area represents binary digit 0. Optical discs are direct access devices. The advantage of optical disks are that the amount of physical disk space needed to record an optical bit is much smaller than usually required by magnetic media and is less susceptible to deterioration.

Cache

It is high-speed memory and expensive. The CPU keeps part of the program and the data into cache memory, which it is likely to use frequently. Cache plays a direct role in increasing Internet speed than RAM does. RAM and cache are both a form of computer memory. Cache stores images and text from web pages as you view them. Pages in cache are not updated while RAM refreshes it again and again.

Traditionally, Cache is categorized as “levels” that describe its closeness and accessibility to the microprocessor. There are three general cache levels: **L1 cache**, or primary cache, is extremely fast but relatively small, and is usually embedded in the processor chip as CPU cache. **L2 cache**, or secondary cache, is often more capacious than L1. L2 cache may be embedded on the CPU, or it can be on a separate chip or coprocessor and have a high-speed alternative system bus connecting the cache and CPU. That way it doesn’t get slowed by traffic on the main system bus. **Level 3 (L3) cache** is specialized memory developed to improve the performance of L1 and L2.

Activity A

1. Computer’s motherboard, processor, and CPU are the same or different things?

.....
.....
.....

2. What’s the difference between cache and RAM?

.....
.....
.....

2.6 INPUT/ OUTPUT DEVICES

They are used as a medium of communication between the external environment and the CPU. The input is provided to the computer through

the input devices and the output is given through the output devices. The input and output devices are controlled by the CPU. A computer system may also have devices under the control of CPU dedicated for controlling the input output devices.

Input Devices

Table 2.2 highlights the categories into which the input devices fall. Voice Recognition devices are used to recognize human speech. It converts human voice signals received through microphone to digital signals using software tools. Voice recognition systems are becoming popular.

Table 2.2: Categories of Input devices

Category	Name of Device
Keying Device	Keyboard
	Punched card reader
Pointing Devices	Mouse
	Touch Screen
	Joy Stick
	Light Pen
Voice Recognizers	
Optical Character Recognizer	Bar Code Reader, QR Code Reader
	Optical Mark Reader
	Optical Character Reader / Optical Scanner
Other Devices	Cameras
	Digitizers (for maps, graphs etc)
	Smart Cards
	Telephone and other handheld devices for capturing data and information

Bar Code Reader scans the code (black and white bars), which specifies the name of the product and its manufacturer. The computer then finds the price of the product from the database. It is used for high- volume processing of data where use of keyboard entry can be very time consuming, for example they are used in supermarkets. QR Codes are used to specify details of data and information contained in a document. It is used for establishing the authenticity of data and information contained in it.

Digital cameras are used to capture pictures. The pictures are digitized and stored in computers. The pictures can be reviewed, deleted, edited, and saved for future use. Images can be transmitted from a PC to a printer or to other cameras. They can be mailed to friends. Digitizers are devices that convert drawings made on sensitized surface with the help of pen to machine-readable input. These inputs are transferred to the computer. A smart card is a card that is embedded with either a microprocessor and a memory chip or only a memory chip with non-programmable logic. The microprocessor card can add, delete, and otherwise manipulate information on the card, while a memory-chip card (for example, pre-paid phone cards) can only undertake a pre-defined operation.

The output generated by a computer can be transmitted to the user via several devices and media. These devices are called output devices. Following are some important output devices:

1. *Monitors:* It is a video screen that displays both input and output data. It comes in different sizes ranging from few inches to several feet. It is interactive in nature, which is the main advantage associated with this device. It consists of CRT (Cathode Ray tube) technology. It is the electronic “gun” that shoots the beam of electrons and illuminates the pixels on the screen.
2. *Printers:* There are two broad categories of printers i.e. Impact and Non Impact Printers. Impact printer uses striking action to press a carbon against paper to create a character. Dot matrix, Line, and Daisy wheel are some important type of Impact printers. They are slow and noisy. They don’t support graphics and are susceptible to mechanical breakdowns.

Non Impact Printers use laser beams to write information on photosensitive drums. The paper passes over the drum and picks the image with toner. Laser Printer is the example of Non Impact Printers. They have high speed and can produce print- quality text and graphics. Ink-jet printers are Non Impact Printers, which shoot tiny dots of ink on the paper. They are inexpensive and are used for low volume graphical applications when different colors of ink are required.

3. *Audio Devices/Speakers/Headphones:* The outgoing audio data requires the sound amplifying devices, and a few such audio output devices are Speakers & Headphones. These output devices are connected to the main computing devices in order to get the output audio data from the processing unit. An important hardware unit is required or enabling a smooth function between the computer system and the audio devices are the Sound card.
4. *Projector:* Projectors are another type of display devices, which are used for displaying the outcomes from the computer system on to a different exterior surface, by reproducing the same computer’s contents for projection. It is a common understanding that the output data can be displayed clean and clear if the display surface is a plain hard surface white in color. This process is made possible using an extra hardware card, known as the Video Card. A very common usage of Projector devices are Movie Theater and Business Presentation.
5. *Plotter:* They use computer driven pens for creating black and white or color graphic images likes charts, graphs etc. They are used in engineering and architectural drawings.
6. *Television:* Televisions are the most famous and commonly used output devices, which are used to display the video content and for playing the audio files. It is a well-liked way of exhibiting the videos, images, presentations, etc. for domestic as well as the technology-related industries. It is generally called as TV or Telly that has had

evolved over the years from CRT screens to LCD displays to LED screens to the latest Plasma displays.

7. *Braille Reader*: Braille Readers, as the name says, are used to facilitating the computer functional handling for blind users. The workings of this device involve the central processing system to deliver the processed data in the Braille format, so as to make it recognizable for the visionless disables. When the Braille Reader receives the processed data, the data is converted to push the pins against a flat surface. When the visually disabled use this device, they can run their fingers through these Braille script and understand the outcome easily. Similar to other output devices, this device comes in a variety of sizes as well.
8. *GPS*: GPS stands for Global Positioning System device. This device uses the satellite resources for capturing the geographical position of the device. Hence, if the user holds this device in his hands the geographical location of the user can be captured accurately. It involves the calculation of exact latitude and longitude points for the location required and share the outcomes to the user. In the evolving technological world, all the portable devices are equipped with GPS systems, as it helps in locating the device when lost or stolen. The devices that come with inbuilt GPS units are mobile phones, wearable digital tracking devices, laptops, tablets, cars, bicycles, watches, and even shoes.
9. *Video Card*: Video Card is a hardware device that can be placed in the dedicated sockets provides inside the motherboard of the computer's processing unit. It is typically used for improvising the appearance of the digital content displayed in the other display based output devices. In recent days, this has become an indispensable element to keep up with the growing quality of content in the latest systems.

2.7 NETWORKING AND NETWORKING DEVICES

A network is a group of computers, printers, and other devices that are connected together with cables. Data travels over the cables, allowing network users to exchange documents and data with each other. Many users can use the same printer and share hardware and software that are connected to the network. Each computer, printer, or other peripheral device that is connected to the network is called a node. A network can have tens, thousands, or even millions of nodes.

The number of possible ways to logically arrange the nodes, or computer systems and devices on the networks are mainly of five types. They are termed as network topologies, i.e. the logical model that describes the network structure or configuration. These types are bus, ring, hierarchical, star and hybrid.

The computers in the ring network are connected in ring or circle. There is no central computer and messages are routed around the ring from one device to another in one direction. In the bus network the devices are connected on the single line. Each device is connected to the single bus

and can directly communicate with all other devices on the network. A star network has a central computer. This central computer controls and directs messages. If the central computer breaks down, entire system is broken down. The hierarchical structure has tree like structure and does not have a central computer. Messages pass through all the branches of the tree till they reach the destination. This structure is easy to repair as any branch can be isolated and repaired without affecting any other branch. Hybrid network is the combination of two or more topologies.

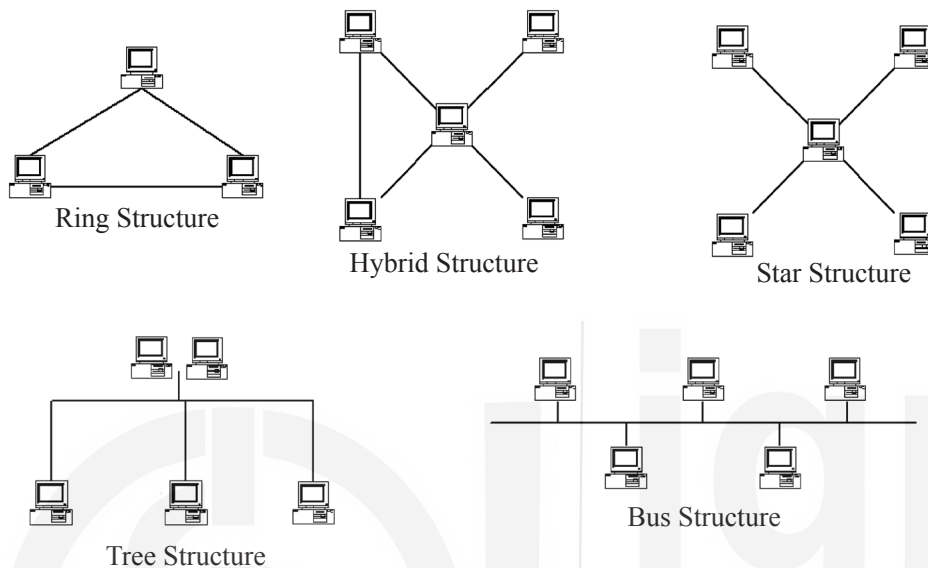


Figure 2.3: Types of Network Topologies

Network Devices

Network devices allow multiple computers to exchange data and information through Internet or Intranet connection. Many people can use a single printer, or share other devices and resources on the network. Every network has a router which is a device that connects two or more networks and determines the next point to which a 'packet' of data should be forwarded towards its destination. A packet is the smallest unit of data that travels on the network. Hub and Switch are the center points where several network connections converge i.e. they are the points where data comes in from one or more directions/locations and is forwarded to one or more directions/locations. Sometimes the same device serves as both the switch and router. NIC (Network Interface Card) is a card installed on a computer that connects the computer to the network. It provides dedicated and full-time connection to the computer.

2.8 PLUG AND PLAY DEVICES

Plug and Play (also called *PnP*) specification was developed by Microsoft with cooperation from Intel and many other hardware manufacturers. The goal of Plug and Play is to create a computer whose hardware and software work together to automatically configure devices and assign resources, to allow for hardware changes and additions without the need for large-scale resource assignment tweaking. As the name suggests, the goal is to be able to just plug in a new device and immediately be able to use it, without complicated setup maneuvers.

The plug and play feature was introduced in Windows 2000. One can plug in, for instance, an USB (Universal Serial Bus) device and it is detected and accepted by the system so that the device can be used without rebooting the system. Operating system has drivers for most of the devices. When a device, for which operating system has the driver, is plugged in, the communication between the driver and the device is established by the OS. These devices are connected to the computer in several ways. Some devices, such as network adapters and sound cards, are connected to expansion slots inside the computer. Other devices, such as printers and scanners, are connected to ports outside the computer. Some devices, known as PC Cards, connect only to PC Card slots on a portable computer. The operation system must have device drivers for all plug and play devices.

USB is the industry standard for attaching peripheral devices to the computer. This technology is designed for use with almost all devices, including printers, digital cameras, game pads, joysticks, keyboards and mouse, TVs, storage devices. USB hardware is identical across platforms. It replaces all the different kinds of serial and parallel port connectors with one standardized plug and port combination. The devices can plug in without opening the PC.

Activity B

What were the main reasons for the change from the serial, printer, game, and ps/2 ports to the USB port?

.....

.....

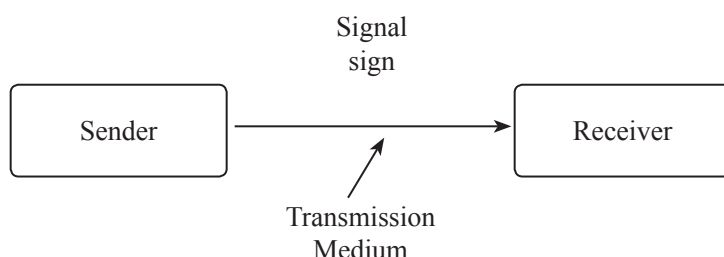
.....

2.9 SMART DEVICES

Smart devices are interactive electronic gadgets that understand simple commands sent by users and help in daily activities. Some of the most commonly used smart devices are smartphones, tablets, phablets, smartwatches, smart glasses and other handheld personal electronics items. With the digitalization in practice almost in every field such devices are very commonly seen everywhere. A smart device is able to connect, share and interact with its user and other smart devices. Although usually small in size, smart devices typically have the computing power of a few gigabytes. Such devices are very common these days in information gathering and further dissemination. They are all of the everyday objects made intelligent with advance computing including artificial intelligent and machine learning. They are networked to form the internet of things (IoT). These small devices are powerful enough to process data. They range from sensors to refrigerators and wearable to container transportation, capable of running autonomous workloads.

2.10 COMMUNICATION TECHNOLOGY

Communications is the transmission of a signal by a way of particular medium from a sender to a receiver. For example in human speech, the sender transmits a signal through the transmission medium air.



For the effective communication both the sender and the receiver should understand the signals and have common interpretation for them. When communication has to be established a medium is required for transmission. This pathway/medium is called as communication media. It includes twisted pair, coaxial cable, fiber optic cable, microwave transmission, and satellite transmission. With fiber-at-home communication media you get high speed internet connectivity at your home with applications like internet TV (iTV) giving the facility of watching TV channels, browsing web sites and can play and view audio/video programs of your interest.

There are two modes of communication – Synchronous and Asynchronous. In Synchronous transmission, data bits are sent over a communication link and the data transfer is controlled by a timing signal initiated by the sending device. Each data block is preceded by unique characters called sync bits. The receiving device synchronizes itself with a stream of these bits. Synchronous transmission is generally used for transmitting large volume of data at high speed. For Asynchronous transmission, one character is transmitted or received at a time. Each character is preceded by start bit and ended with end bit, so that the receiving device knows where the character begins and ends. It is inefficient due to the overhead of transmitting start and stop bits. It is generally used for low speed data transmission.

Data transmission occurs in one of the following three directions: Simplex, half duplex, or full duplex. In simplex data transmission, the data transmission takes place in only one direction. Public announcement systems such as radio and television use simplex mode of data transmission. It is simple and relatively inexpensive. In half- duplex transmission mode, the transmission takes place in both directions but only in one direction at a time. Full duplex mode of transmission allows transmission in both directions simultaneously. For example, on telephone both parties can communicate simultaneously. The cost of full duplex transmission is high as compared to both simplex and half duplex.

2.11 TYPES OF COMPUTERS

Mainframes: Mainframes are not as powerful and not as expensive as super computers. A large organization generally uses a mainframe as a contact database and file server.

Minicomputers: There are also called midrange computers. There are less powerful than mainframes. Minicomputers are used for specific tasks such as scientific research and engineering applications. Some times, large organizations install many minicomputers instead of one mainframe. The application, and data are distributed ones minicomputer to achieve better reliability. Small organizations use a minicomputer as a server.

Workstations: A workstation is usually based RISC architecture, provides high resolution graphic displays and high-speed calculations. Scientific and business community uses these machines.

Microcomputer: Microcomputers, also called personal computers, are general-purpose computers. A PC that sits on a desk is called desktop personal computer in contrast to a laptop computer, which is portable.

Network Computer: A network computer (NC) is a desktop terminal that is connected to a central computer. A NC has essentially no storage, no CPU, and no secondary storage. It has a keyboard and a video display. It uses processing power and storage of its central processor.

2.12 SUMMARY

In this unit, we have discussed an introduction to computer systems. The components of computer, which we explained in detail include: CPU, ALU, storage, and I/O devices. The CPU works as the brain of the system, while RAM is the working memory of the system. Today's computers come with possibilities to use many plug and play devices, and they are also network compatible. There are smart devices available today for information gathering, processing and further disseminating. We have also discussed the concepts related to networking in this unit, where we have described the various media used such as co-axial cable, satellite, fiber optics, and microwave communication. At the end, we have given a brief overview of different types of computers through generations.

2.13 UNIT END EXERCISES

1. What are the components of a computer system?
2. What are the components of a CPU?
3. What factors affect the speed of a computer?
4. Write a detailed note on primary memory.
5. Write a detailed note on secondary memory.
6. Write a detailed note on Input-Output devices.
7. What are plug-and-play devices?
8. What are various types of computer?

2.14 REFERENCES AND SUGGESTED FURTHER READINGS

Fitzgerald, J. and Dennis A, *Business Data Communications and Networking*, John Wiley & Sons, Inc.

Laudon C K & Laudon J P, *Management Information Systems*, Pearson Education, Asia.

Turban, T., Ephraim, M. and Wetherbe J., *Information Technology for Management*, John Wiley and Sons, Inc.

www.arm.com, site accessed on November 15, 2021.

www.diskpart.com, site accessed on November 15, 2021.

UNIT 3: COMPUTER SOFTWARE

Structure

- 3.1 Introduction
- 3.2 Objectives
- 3.3 System Software
 - 3.3.1 Operating Systems
 - 3.3.2 Language Translators
 - 3.3.3 Utility Programs
- 3.4 Application Software
 - 3.4.1 Programming Languages
- 3.5 Difference between Application Software and System Software
- 3.6 Open Source Software
- 3.7 Acquiring application software
- 3.8 Summary
- 3.9 Unit End Exercises
- 3.10 References and Suggested Further Readings

3.1 INTRODUCTION

The word **software** collectively refers to various kinds of programs used to operate computers and related devices. A **program** is a sequence of instructions that a computer can interpret and execute. Programs can be built into the hardware itself, or they may exist independently in a form known as software. Hardware describes the physical components of computers and related devices.

Software may be distributed on floppy disks, CD-ROMs, and on the Internet. It is usually stored on an external long-term memory device, such as a hard drive or magnetic diskette. When the program is in use, the computer reads it from the storage device and temporarily places the instructions in random access memory (RAM). The process of fetching and then performing the instructions is called “running,” or “executing,” a program. Software programs and procedures that are permanently stored in a computer’s read-only memory (ROM) are called firmware.

The two main types of software are system software and application software. Application software consists of programs that are aimed to help users in solving particular computing problems. Microsoft Edge for web browsing, Adobe Photoshop for developing computer graphics, social media tools like WhatsApp, Facebook, Twitter and Messenger for instant messaging all lies in the application software category. The other class of software is the system software, which encompasses the programs that heavily interact with computer resources and provide services to other programs. Popular examples in this are Operating Systems (OS), hardware drivers, compiler etc.

3.2 OBJECTIVES

After reading this unit, you should be able to

- Describe about the different types of operating systems and their functions;
- State the characteristics of system software and application software;
- Differentiate between compiler and interpreters;
- Differentiate between application software and system software;
- Enumerate the advantages and disadvantages of compilers and interpreters;
- Define and explain the concepts and philosophy of open-source software; and
- Explain the process of software acquisition.

3.3 SYSTEM SOFTWARE

System software co-ordinates the various parts of computer system and mediates between the application software and computer hardware. Operating system is system software, which manages and controls the computers activities. The other system software consists of computer language translation programs that convert programming languages into machine language and utility programs that perform common processing tasks.

3.3.1 Operating System

An **operating system** is a set of computer programs that controls the computer hardware and acts as an interface with the application programs. The operating system plays a central role in the functioning of a computer system. It is usually stored on disk, after the computer system is started or booted up portions of operating system are transferred to memory as required. The **kernel** as the name suggests is the heart of the operating system and controls the most critical processes. Windows by Microsoft, Linux, UNIX, and the Macintosh are the commonly used operating systems.

In some specialized or embedded computers the operating instructions are contained in their circuitry; common examples are the microcomputers found in calculators, automobile engines, mobile phones and microwave ovens.

Functions of Operating System

An operating system performs allocation and assignment of system resources, schedules the use of computer resources, monitors the computer system activities etc. The various activities performed by a typical operating system are:

- Performing common computer hardware functions.
- Providing a user interface
- Providing a degree of hardware independence
- Managing system memory

- Managing processing tasks
- Providing networking capability
- Controlling access to system resources
- Managing files

Common Hardware Functions

All application programs must perform certain tasks. For example

- Getting input from the keyboard or some other input devices
- Retrieving data from disks
- Storing data on disks
- Displaying information on a monitor or printer

Each of these basic functions requires a more detailed set of instructions to complete. The operating system converts a simple, basic instruction into the set of detailed instructions required by the hardware. In effect, the operating system acts as intermediary between the application program and the hardware. The typical OS performs hundreds of such functions, each of which is translated into one or more instructions for the hardware. The OS notifies the user if input/output devices need attention, if an error has occurred, or if anything abnormal has happened in the system.

User Interface

One of the most important functions of any operating system is providing a user interface. A user interface allows individuals to access and command the computer system. The first user interfaces for mainframe and personal computer systems were command based. A command-based user interface requires that text commands be given to the computer to perform basic activities. For example, the command ERASE TAXRTN would cause the computer to erase or delete a file called TAXRTN. RENAME and COPY are other examples of commands used to rename files and copy files from one location to another. Many mainframe computers use a command-based user interface. In some cases, a specific Job Control Language (JCL) is used to control how jobs or tasks are to be run on the computer system.

A graphical user interface (GUI) uses pictures (called icons) and menus displayed on screen to send commands to the computer system. Many people find that GUIs are easier to use because user intuitively grasp the functions. Today, the most widely used graphical user interface is Windows by Microsoft. Alan Kay and others at Xerox PARC (Palo Alto Research Center, located in California) were pioneers in investigating the use of overlapping windows and icons as an interface. As the name suggests, Windows is based on the use of a window, or a portion of the display screen dedicated to a specific application. The screen can display several windows at once. Another example which is also widely used is iOS by Apple. The use of GUIs has contributed greatly to the increased use of computers because users no longer need to know command-line syntax to accomplish tasks.

Hardware Independence

The applications make use of the operating system by making requests for services through a defined application program interface (API).

Programmers can use APIs to create application software without having to understand the inner workings of the operating system.

Suppose a computer manufacturer designs new hardware that can operate much faster than before. If the same operating system for which an application was developed can run on the new hardware, minimal (or no) changes are needed to the application to enable it to run on the new hardware. If APIs did not exist, the application software developers might have to completely rewrite the application program to take advantage of the new, faster hardware.

Memory Management

The purpose of memory management is to control how memory is accessed and to maximize available memory and storage. The memory management feature of many operating systems allows the computer to execute program instructions effectively and with speed.

Memory controller allows the computer system to efficiently and effectively store and retrieve data and instructions and to supply them to the CPU. Memory management programs convert a user's request for data or instructions (called a logical view of the data) to a physical location where the data or instructions are stored. A computer understands only the physical view of data—that is, the specific location of the data in storage or memory and the techniques needed to access it. Memory controller converts a logical address to a physical address.

Memory management is important because memory can be divided into different segments or areas. Some computer chips provide “rings” of protection. An operating system can use one or more of these rings to make sure the application programs do not penetrate an area of memory and disrupt the functioning of the operating system, which could cause the computer system to crash. Memory management features of today's operating system are needed to make sure that application programs can get the most from available memory without interfering with other important functions of the operating system or with other application programs.

Most operating systems support **virtual memory**, which allocates space on the hard disk to supplement the immediate, functional memory capacity of RAM. Virtual Memory works by swapping programs or parts of programs between memory and one or more disk drives using a concept call **paging**. This reduces CPU idle time and increases the number of jobs that can run in a given time span.

Processing Task

Managing all processing activities is accomplished by the task management features of operating systems. Task management allocates computer resources to make the best use of system assets. Task management software may permit one user to run several programs or tasks at the same time (multitasking) and allow several users to use the same computer at the same time (time-sharing).

An operating system with multitasking capabilities allows a user to run more than one application at the same time. Without exiting a program, user

may start another application and switch to newly started application, and then jump back to the first program, picking up where it was left off. Better still, while user is working in the *foreground* on one program, one or more other applications can be churning away, unseen, in the *background*, sorting a database, printing a document, or performing other lengthy operations. In the absence of starting background processes, a foreground process would monopolize the computer and if the process happens to be non-interactive such as a file print, then the user will just have to sit and stare at the screen till the process completed. Multitasking can save users a considerable amount of time and effort.

Time-sharing allows more than one person to use a computer system at the same time. For example, 15 customer service representatives may be entering sales data into a computer system for a mail-order company at the same time. In another case, thousands of people may be simultaneously using an on-line computer service to get stock quotes and valuable business news.

Time-sharing works by dividing time into small CPU processing time slices, which can be a few milliseconds or less in duration. During a time slice, some tasks for the first user are done. The computer then goes from that user to the next. During the next time slice, some tasks for the next user are completed. This process continues through each user and cycles back to the first user. Because the CPU processing time slices are small, it appears that all jobs or tasks for all users are being completed at the same time. In reality, each user is sharing the time of the computer with other users.

The ability of a computer to handle an increasing number of concurrent users smoothly is called scalability. This is a critical feature for systems expected to handle a large numbers of users such as a mainframe computer or a Web server. Because personal computer operating systems usually are orientated toward single users, the management of multiple-user tasks often is not needed.

Networking Capability

The operating system can provide features and capabilities that aid users in connecting to a computer network. For example, Apple computer users have built-in network access through AppleShare feature, and the Microsoft Windows operating systems come with the capability to link users to the Internet.

Access to System Resources

Computers often handle sensitive data that can be accessed over networks. The operating system needs to provide a high level of security against unauthorized access to the users' data and programs. Typically, the operating system establishes a log-on procedure that requires users to enter an identification code and a matching password. If the identification code is invalid or if password does not go with identification code, the user cannot gain access to the computer. The operating system also requires that user passwords be changed frequently. If a user is successful in logging onto the system, the operating system records the details of the user and system

usage. In some, organizations, these records are also used to bill users for system and resource usage. The operating system also reports any attempted breaches of security.

File Management

An operating system performs file management functions to ensure that the files are available to CPU when needed and that they are protected from access by unauthorized users. Many computers support multiple users who store files on centrally located disks or tape drives. The operating system must be able to resolve what to do if more than once user requests access to the same file at the same time. Even on stand-alone personal computers with only one user, file management is needed to keep track of where files are located, what size they are, when they were created, and who created them.

3.3.2 Language Translators

The CPU (also called processor) of a computer understands commands in machine language, where each instruction is a series of binary digits. Programming in machine language is not easy, as programmers have to remember the machine codes, which are in binary format. To help programmers, other high level programming languages have been developed whose instructions are easy to remember for programmers as these languages use English words. C, Java, SQL are examples of high level programming languages. Programming languages can be divided into assembly languages and high-level programming languages.

For any program to be executed, it has to be first converted into its equivalent machine language program and then loaded into the memory of computer. To perform the translations of programs, language translators are used. As the process of programming language translations are machine dependent, the translators fall in the system software category.

Assemblers: The computer software that translates the assembly language programs into corresponding machine language programs are known as *assemblers*. Assembly language uses mnemonics instead of binary codes used in machine language. For example ADD R1 R2 is an assembly language instruction for adding the contents of register R1 with the contents of register R2 and store the result in R1. The use of mnemonics helps programmers to remember programming codes. But still to write big programs like a word processing software can be very cumbersome in assembly language.

Compiler and Interpreter: Compiler and interpreter are used to translate a high level programming language program into a machine language program. As the translation process is very cumbersome, some compilers first translate the source code (the program in high-level language) into the equivalent assembly language program and then use the assemblers for the next step. To define, a compiler is a program that translates a source text written in a language A into a target program in language B, whereas, interpreter is a program which directly executes the program in a given programming language A.

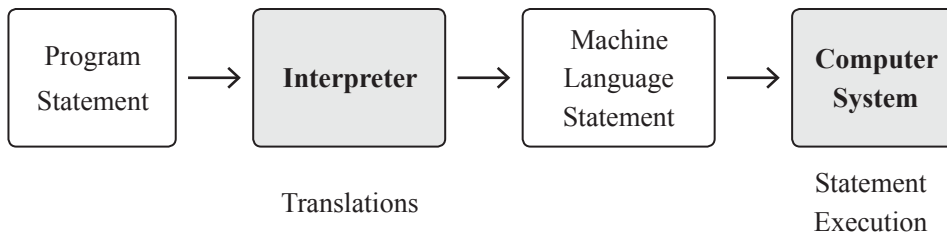


Fig 3.1: Working of Interpreter

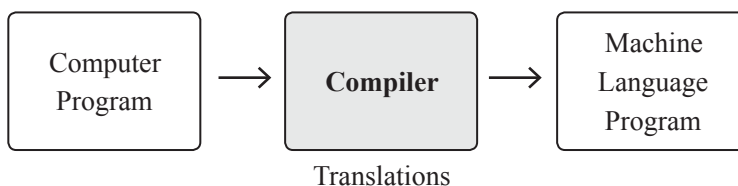
Difference between Compiler and Interpreter

An interpreter is a language translator that reads the source code line-by-line and executes them one by one. On the other hand, a compiler first reads the complete source code and then generates its object code (the equivalent machine language program).

Interpreter has the advantage that the process of translation takes less time as compared to the compiler, but the program generated is less efficient as compared to the compiler in terms of the time it takes to execute the program.

Compilers are generally used when the efficiency of the generated binary (machine language) program is desired. Interpreters are mostly used for educational purposes, where the programmer makes frequent translations of the program as the changes to the program are frequent. Also, as the interpreter generates the binary program on the fly, platform independence can be achieved. Platform independence means the capability of the program to be translated on any platform and executed on any platform. By platform we mean the hardware and the operating system running on that system. For example, interpreted languages such as HTML, Perl, and Lisp are platform independent. Java, which is a half-compiled half-interpreted language, has benefits of both. As Java programs are compiled to byte-code, instead of binary code, it can be ported to any platform which has the Java Virtual Machine software, which interprets the byte-code. By compiling the program to byte code certain level of optimization that are possible in compilers are achieved.

Step 1: Program Translation



Step 2: Program Execution

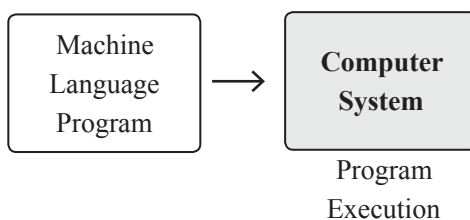


Fig 3.2: Working of Compiler

Advantages and Disadvantages of Interpreter over Compiler

Advantages:

- As compared to compiler no synthesis phase is required in interpreter: Neither there is a need to learn target language B nor the target code is to be generated. Thus interpreters don't have synthesis phase.
- Direct Execution: There is no intermediate compilation phase so the code is directly executed.

Disadvantages:

- Efficiency Loss: As the code is executed on the fly, the efficiency of the program is low. In compilers, there is a separate phase for optimization of the program code.
- Interpreter must be available on target machine: The compiled code can be executed on any similar machine. The code needs not to be compiled every time. For languages, which are interpreter based, the interpreter must be available on each machine where the code is to be executed.

Activity A

Why is there a difference between High-Level and Low level languages?

.....

.....

.....

.....

.....

3.3.3 Utility Programs

A utility program is designed for general support to the processes of a computer. They are usually for routine, repetitive tasks and many users share them. Examples of utility programs include diagnostic programs, antivirus tools, trace programs, input routines, and programs used to perform routine tasks, *i.e.*, perform everyday tasks, such as copying data from one storage location to another. Utility programs are also available commercially; for example, Norton Utilities package is a set of utility programs for checking disks for computer viruses, checking hard drive for bad locations and removing them and for performing disk compression. Some of the examples of other such software for checking the health of hard disks are SeaTools by Seagate, HDDScan, Fujitsu Diagnostic Tool.

3.4 APPLICATION SOFTWARE

Application software is a complete, self-contained program that performs a specific function directly for the user. This is in contrast to system software, which exists to support application programs. Application software may consist of a single program, such as an image viewer; a small collection of

programs (often called a software package) that work together to accomplish a task, such as a spreadsheet or text processing system; a larger collection (often called a software suite) of related but independent programs and packages that have a common user interface or shared data format, such as Microsoft Office, which consists of closely integrated word processor, spreadsheet, database manager, etc.; or a software system, such as a database management system, which is a collection of fundamental programs that may provide some service to a variety of other independent applications. Some of the examples of application software according to their types are given below in the table 3.1.

Table 3.1: Examples of Application Software

Type	Software
Word Processing	Microsoft Word, GoogleDoc, WordPerfect
Spreadsheet	Microsoft Excel, Apple Numbers, Google Sheets, Quip, Smartsheet
Graphics	Apple Pie Design, Adobe Photoshop, Corel Draw, Microsoft Paint
Desk Top Publishing	FlipHTML5, Quark Express, Adobe InDesign, Microsoft Publisher, LibreOffice Draw
Social Media messaging	Facebook, LinkedIn, WhatsApp, Twitter
Virtual Conferences	MilanSetu, Zoom, Google Meet, Skype, Cisco Webx,

Activity B

Explain the difference between Applications and Utilities with the help of examples.

.....

.....

.....

.....

3.4.1 Programming Languages

A programming language is an artificial language (as opposed to natural languages such as Hindi, English etc.) that is used to generate or to express computer programs. Both system software and application software are developed using one or many programming languages.

Generations of Programming Languages

The programming languages have been divided into different generations according to their characteristics and capabilities.

- **Machine Language** The first generation of computer programming languages is machine language. Programs in machine language consist of instructions coded in of 0s and 1s, thus the alphabets of machine language are 0 and 1. The storage locations and data items are also specified using 0s and 1s. These languages are machine dependent.

There is a machine language corresponding to each microprocessor available.

- **Assemble Language** The second generation of computer programming language started using mnemonics (like ADD, SUB) to represent machine language instructions and storage locations. Assembly language is also machine-dependent. System software or at least part of it is usually developed in assembly languages.
- **Third Generation Language (3GL)** 3GL are English-like languages. They use statements and commands, which are similar to the words used in English. 3GLs are easier to learn, but less efficient in the use of computer resources as compared to machine and assembly languages. Typically, a statement in 3GL is translated into many instructions of machine language. C, BASIC, FORTRAN, COBOL and Pascal are the popular third generation languages.

FORTRAN (FORMula TRANslation): Fortran was developed in 1956 by John Backus. It was developed keeping in mind the scientific and engineering application.

COBOL (Common Business Oriented Language): COBOL was developed in the early 1960s under the auspices of the U.S. Department of Defense in cooperation with computer manufactures, users, and universities. It was designed to be a language for writing programs for business problems. Another design objective was to keep it machine independent. The language was designed in such a manner that it could evolve and grow to take care of changing program development requirements. Many standards for COBOL have been published since then.

BASIC (Beginners All-purpose Symbolic Instruction Code): BASIC was developed in 1964 by John Kemeny and Thomas Kurtz to teach students at Dartmouth College to use computers. It was meant to be a very simple language to learn and also one that would be easy to translate. Furthermore, the designers wished it to be a foundation language for students who wished to learn more powerful languages such as FORTRAN or ALGOL.

Pascal: It was developed in the late 1960s by Niklaus Wirth of Zurich. He named it after the great mathematician and philosopher, Blaise Pascal. Both Pascal and BASIC have been used extensively for teaching to the beginners.

C: It was developed at Bell Laboratories in 1972 by Dennis Ritchie. Many of its principles and ideas were taken from the earlier language B and B's earlier ancestors BCPL and CPL. C was developed with the purpose of creating a high level language that could be used for writing machine independent programs and would still allow the programmer to control the behavior of individual bits of data. The bit processing features have made C a popular system software development language.

- **Fourth Generation Languages (4GL):** Fourth generation languages are less procedural and even more English-like than 3GL languages. The emphasis is more on the output format than the procedure applied to achieve the results. That is why 4th generation languages are also called as non-procedural languages. Some 4GLs are given in Table 3.2.

Table 3.2: Fourth Generation Languages

4GLs	Salient Feature
Natural, Power Builder, Visual C++, Visual Basic	Application Generators
SQL, RPG-III	Query Languages/ Report Generators
Systat, SAS Graph	Graphics Languages
Oracle HRMS, Oracle Financials, CRM, PMS, SAP Modules	Application Software Packages

- **Object Oriented Languages:** Object Oriented Programming is a type of programming in which programmers define the data types and structure of the data, in addition, the operations (functions) that can be applied to the data is also defined. The data and functions together are known as objects. Programmers can create relationships between objects. For example, objects can inherit characteristics from other objects. One of the principal advantages of object-oriented programming techniques over procedural programming techniques is that they enable programmers to create modules that do not need to be changed when a new type of object is added. A programmer can simply create a new object that inherits many of its features from existing objects. This makes object-oriented programs easier to modify. The object-oriented programming languages are usually referred to by the acronym OOPL, Java, C++ and Smalltalk are the popular languages.
- **Python:** It was developed by Guido van Rossum in the late eighties and early nineties at the National Research Institute for Mathematics and Computer Science in the Netherlands. Python is derived from many other languages, including ABC, Modula, C, C++, Algol, SmallTalk, and Unix shell and other scripting languages. There is a fact behind choosing the name Python. Guido van Rossum was reading the script of a popular BBC comedy series “**Monty Python’s Flying Circus**”. It was late on-air 1970s. Van Rossum wanted to select a name which unique, sort, and little-bit mysterious. So he decided to select naming Python after the “Monty Python’s Flying Circus” for their newly created programming language.

3.5 DIFFERENCE BETWEEN APPLICATION SOFTWARE AND SYSTEM SOFTWARE

System software and application software are computer programs. The system software is also installed during the installation of the operating system. However, the application software utilizes the capabilities of the computer on which it is installed.

The programs and the file that comprises the operating system are system software. These files include configuration files, system preferences, system services, libraries of functions and the drivers for the hardware installed on the computer. The computer programs in system software include compilers, system utilities, assemblers, debuggers and file management tools.

Once you install the operating system, the system software is also installed. Program such as “Software update” or “Windows update” can be used to update the system software. However, the end user does not run the system software. For example, while using the web browser, you don’t need to use the assembler program.

System software is also called low-level software as it runs at most basic level of the computer. It just creates a graphical user interface thorough which the user can interact with hardware with the help of operating system. System software just runs at the back so you don’t need to bother about it.

The system software provides an environment to run application software and it controls the computer as well as the applications installed on the machine.

Application software utilizes the capabilities of computer. When multiple applications are packaged together then it is called application suite.

There is a common user interface in each application suite which makes it easier for the user to learn different applications. In some cases, such as Microsoft Office, the various application programs have the ability to interact with each other. This facility is very handy for the user. For example, a user can embed the spreadsheet in a word processor using the application software. Application software cannot run without the presence of system software.

Now, let us see the difference between system software and application software. It can be listed as below -

1. System software gets installed when the operating system is installed on the computer while application software is installed according to the requirements of the user.
2. System software includes programs such as compilers, interpreters, drivers, assemblers while application software includes software to automate business processes like ERP, CRM, SCM, PMS, and other such software developed by organizations for automation. Media players, social media Apps, word processors, spreadsheet programs are application software.
3. Generally, users do not interact with system software as it works in the background whereas users interact with application software while doing different activities.
4. A computer may not require more than one type of system software while there may be a number of application software programs installed on the computer at the same time.
5. System software can run independently of the application software while application cannot run without the presence of the system software.

3.6 OPEN SOURCE SOFTWARE

Open Source Software (OSS) is primarily defined as software, which is freely re-distributable and includes the source code. The licenses under which OSS is released vary greatly. They can be easily downloaded from

websites. OSS is vastly different from the mainstream software industry where source code is highly guarded and programs are only distributed in non-modifiable format.

The most important aspect of the open source movement is the participation of users. When a user wants a feature or a bug fix for a commercial program, the user is at the mercy of the software vendor. However, with open source, the user can modify the program according to his needs or fix a bug. Many users will help develop the program for free, simply to improve the product and for the benefit of the community.

These are a few of the most common and popular licenses for OSS. Some examples are -

1. GNU Public License (GPL) V2 & V3
2. GNU Affero General Public License (AGPL)
3. Limited GNU Public License (LGPL)
4. MIT License
5. Berkeley Software Distribution (BSD) 3-Clause License
6. Apache License 2.0
7. The Artistic License
8. The Netscape Public License (NPL)
9. Mozilla Public License (MPL)
10. Apple Public Source License (APSL)

A few advantages of OSS are:

- 1) **Cost Effective:** Open source software often comes free. The individual or organization users can save the software cost.
- 2) **Customizable:** Since Open source software comes with the source; one can customize existing software to suit one's needs. Closed source software *may* be customizable, but you need to negotiate and/or pay for customization. Open source licenses typically *guarantee* you the right to be able to customize the software.
- 3) **More Secure:** Since the source code is open, more people scrutinize the source code, and hence more flaws are found and corrected. The end result is that the code produced is more secure compared to similar closed-source code.

The following terms are synonymous with Open Source Software: Freeware, Free/Libre/Open Source Software (FLOSS), and Free Open Source Software (FOSS).

Activity C

Explain the differences between commercial software, shareware, open source software, freeware, and public domain software.

.....

.....

.....

.....

3.7 ACQUIRING APPLICATION SOFTWARE

A company can either develop or purchase software for its use. In some cases, the purchased software can be modified / customized according to the needs of the company. The different options available are summarized in the diagram below.

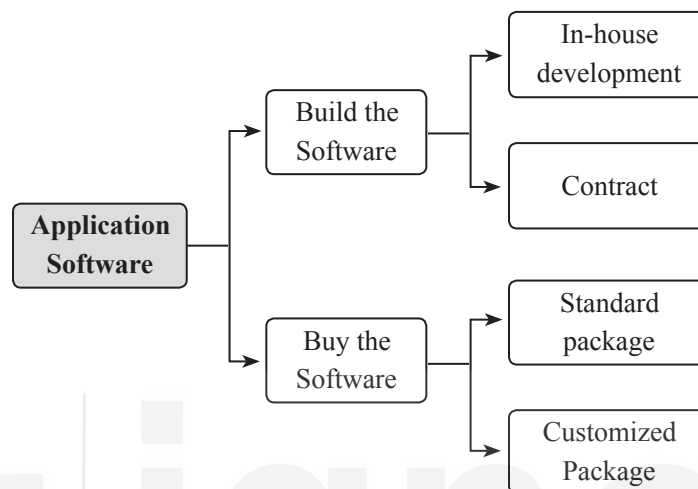


Fig 3.3: Sources of Acquiring Application Software

Build the software If the requirements of the company are unique or specific, then the decision to build the software may be taken. If the organization has the required talent and time, it may be built by the company itself. This kind of development is known as *in-house development*. Also the company may obtain customized software from software vendors. Such software developed for particular companies are called *contract software*.

Buy the software The Company has another option of purchasing, leasing, or renting software from software companies, who develop programs and sell them to many computer users and organizations. The software developed for the general market is called *off-the-shelf software*. They are readily available and many companies use them to support their business processes.

Customized Software The Company can also opt to go for a mix of both buy and build decision. In that case, the company can purchase some off-the-shelf available software, and customize it to its needs by in-house or external personnel. There are software vendors in the market who provide a range of services like installing, modifying software, training end users, etc. They can be contracted to do the customization.

3.8 SUMMARY

Computer software have developed so much over the past years that it is very difficult to cover all aspects of the same. System software and application software represents two broad levels of categorization. System software encompasses of the operating system, language translators, and the utility programs. Application software is aimed to solve particular user computing problems. Open source software is distributed with the source code and freely available at a fraction of cost as compared to proprietary

software. Acquiring application software is an important business activity and requires to be managed carefully.

3.9 UNIT END EXERCISES

1. What are the main types of software?
2. What are the main functions of operating system?
3. Define multi-tasking and time-sharing system?
4. Difference between compiler and interpreter?
5. Name different generations of programming languages and their characteristics?
6. What are the advantages of open-source software?
7. Describe the decision making process of acquiring application software?

3.10 REFERENCES AND SUGGESTED FURTHER READINGS

Laudon C K & Laudon J P, *Management Information Systems*, Pearson education, Asia.

Stair R, Reynolds G, *Principles of Information Systems; Course Technology*.

Turban, T., Ephraim, M. and Wetherbe J., *Information Technology for Management*, John Wiley and Sons, Inc.

<http://www.differencebetween.com>, site accessed on November 16, 2021.

UNIT 4: NETWORKING TECHNOLOGY

Structure

- 4.1 Introduction
- 4.2 Objectives
- 4.3 Components of a Network
- 4.4 Types of Networks
- 4.5 Advantages of using Network and Internet
- 4.6 The Layered Architectures and Communication Protocol
- 4.7 Client/server architecture
- 4.8 Cloud Architecture
- 4.9 Summary
- 4.10 Self-Assessment Exercises
- 4.11 References and Suggested Further Readings

4.1 INTRODUCTION

There are many networks in the world, some are natural, and some are man-made. There is a network of veins and arteries in the body and there is a network of railroads. But these days when we say networks, we mean data communication networks. Data communication is the movement of computer data from one computer to another. The data may travel as an electrical or an optical signal on transmission systems. Such systems are called data communication networks or simply network. The networking helps us in getting the information fast. If we look around, we can identify many applications that use networks. One of the prime examples is the railway reservation system. This system uses networks to find out the availability of a seat for a passenger. The passenger may be at station A and he can make a reservation from station B to station C. Earlier system of reservation has become history. Networks and advancements in getting fast connectivity have changed the life for better.

4.2 OBJECTIVES

After reading this unit, you should be able to

- Identify the components of a network;
- Describe different types of data communication networks;
- Distinguish LAN, MAN, SWAN and WAN, and Internet and Intranet;
- State the advantages of Internet and Intranet;
- Identify the seven layers of the OSI model of network architecture;
- Explain the TCP/IP communication protocol;
- Discuss benefits to an organization due to networks; and
- Describe the client/server model of computing.

4.3 COMPONENTS OF A NETWORK

A network system consists of hardware components as well as software components.

The hardware components are the following:

The Server or Host Computer - Host is the computer that has data to be transmitted.

The Client - This is the computer on the other end of the transmission system as the server. It receives the transmitted data from the server.

The Network Interface Card: Earlier, one had to buy a network interface card (NIC) separately and install it in the computer. Nowadays, the NIC has become a standard component of a system. The NIC is the interface between the network cable and the computer.

The Circuit: The circuit is the pathway through which data travels from the host to the client. The circuit may be a copper wire or an optical fiber. Earlier the commonly used media were twisted-pair cables and coaxial cables. These days, mostly fiber optic cables are also being used. Fibre at home has taken over the reign. Fiber optic cables can withstand higher temperature and has much higher bandwidth. Microwaves are also used for data transmission. The communication may take place through wireless medium.

Network Hubs - A network hub is a node that broadcasts data to every computer or Ethernet-based device connected to it. A hub is less sophisticated and cheaper than a switch. Network hubs are best suited for small, simple local area network environments. Hubs, however, cannot provide routing capabilities or other advanced network services. Because they operate by forwarding packets across all ports indiscriminately, network hubs are sometimes referred to as “dumb switches.” Hubs support half-duplex transmission i.e. only one device connected to a port can send or receive data at a time but not both at the same time. It transmits data from one device to another in form of binary bits. It may have 4 or multiple of 4 ports, generally up to 16 ports. Each port may have one device connected. It does not store any MAC/IP address to transfer data.

Switch - A switch is a multicast networking device that connects a bunch of computers or devices in a network. It's mainly used to send a private message and it does not waste data. A switch can easily identify which device is connected to which port by using a MAC address giving it the ability to deliver the message to a particular machine because it stores and uses the MAC address of a device to transfer data. It is commonly used in local area networks for connecting many nodes. Unlike Hubs, switches work in full-duplex mode of data transmission because a device connected to a switch can send and receive data at the same time. A switch divides a LAN into many domains, each with its broadband connection, considerably improving the LAN's bandwidth thus giving faster speed as compared to hubs.

Router - a router is a networking device that operates to connect two or more networks. It is a device that establishes a common link between networks to enable data flow between them. It uses the IP address of the device to transfer data between networks. They define the direction in which data flows between two communicating devices. Like switches, routers support full-duplex transmission. Thus, a bunch of computers can send data at the same time. It is commonly used in Local Area Network, Metropolitan Area

Network (MAN) and Wide Area Network (WAN). It manages traffic by forwarding data packets to their proper IP addresses. Traffic between these networks may be managed. It determines the best path to send packets. Routers provide connections between different network architectures. They perform the traffic directing functions on the Internet. Data sent through the internet, such as a web page or email, is in the form of data packets. A packet is typically forwarded from one router to another router through the networks that constitute an internetwork (e.g. the Internet) until it reaches its destination node.

Firewall is a network security system that monitors and controls incoming and outgoing network traffic based on predetermined security rules. A firewall typically establishes a barrier between a trusted network and an untrusted network, such as the Internet.

In the above paragraphs, you have seen new terms like IP address, LAN, MAN, WAN. These terms have been explained further in other sections of this Unit.

Network Operating System - On top of the hardware, there is always special purpose network software, which makes the hardware work. The Network Operating System (NOS) is the software that controls the network. NOS have software for the server as well as for the client. They connect the multiple devices and computers on the network and allow them to share resources on the network.

Following are few examples of network operating system -

- Microsoft Windows Server
- UNIX/Linux
- SONIC
- Cisco IOS
- Artisoft's LANtastic

Cisco IOS is a family of network operating systems used on most Cisco Systems routers and current Cisco network switches. Earlier switches ran the Catalyst Operating System CatOS.

SONIC is a Linux based network operating system developed by Microsoft. It is free and open source NOS.

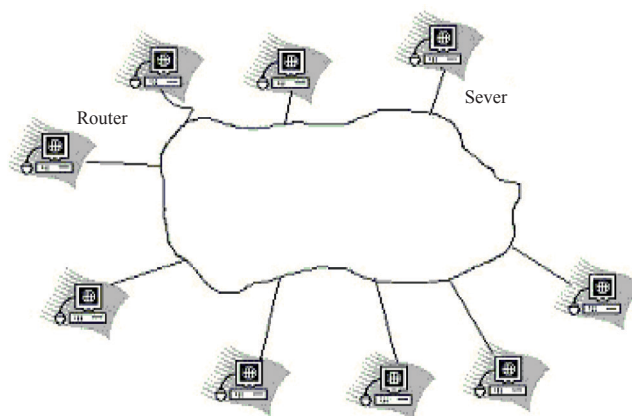


Figure 4.1: A network

Activity A

Identify 5 services around you that use data communication networks.

.....

.....

.....

.....

.....

.....

.....

4.4 TYPES OF NETWORKS

Local Area Network

A local area network connects computers that are in the same building. A network spread over few kilometers also comes under LAN. The LAN is usually realized using Ethernet technology or token ring technology. Fiber Distributed Data Interface (FDDI) is also becoming popular. The transmission rate varies from 10 Million bits per second to 1-giga bits per second (10Mbps-1Gbps). A computer has a network interface card such as Ethernet card that connects it to the network circuit. Usually LAN is under the control of a single organization. The prime objective of LAN is to facilitate information and resource sharing within an organization. For instance, a application software which is used by many people in the organization can be installed on a computer. This computer is connected to other computer by LAN and everybody can use the same software. The machine on which the software is installed is often called a server. In the absence of LAN, the same software will have to be installed on all machines, which may be very expensive. In addition, an upgrade in the software will require re-installation/modification on all computers.

The server may be a file server, print server or a database server, depending on the service it provides to its users. LAN is also connected to other LANs through a gateway. Almost seventy percent of all LANs in the world use Ethernet. Ethernet uses a bus topology. All computers are connected to one circuit. All messages from a computer flow on to the central cable and through it to all computers on the LAN. In other words, messages are broadcasted.

Metropolitan Area Network

A metropolitan area network (MAN) is a computer network that is larger than a single building local area network (LAN) but is located in a single geographic area that is smaller than a wide area network (WAN). Generally, it is a network system of several LANs interconnected by dedicated backbone connections. It may also refer to public use networking infrastructure in a municipality or region. Traditionally it refers to a private data network used by a single organization in several buildings or by several organizations interconnected in the same geographic vicinity. It is larger than a LAN in

a single building but not large enough to be considered as WAN. The size usually ranges from 5 kilometers to 50 km. If all the buildings are on a single piece of contiguous property, it may also be considered a campus network. The backbone connections can use a variety of link technologies, including Ethernet, optical fiber, point-to-point Wi-Fi, Wireless LAN (WLAN), and microwave radio links.

State Wide Area Network (SWAN)

SWAN is the concept of connecting State wide offices at the levels of State/Union Territory Headquarters up to the Block level via District/ sub-Divisional Headquarters, in a vertical hierarchical structure. SWAN is the converged backbone network for data, voice and video communications throughout a State/UT with the following salient features:

- One PoP (Point of Presence) at each State/ District/ Block Headquarter,
- Each PoP has Configurable Aggregation Equipment to enable vertical & horizontal connectivity Gateway to NICNET (National Backbone) for Inter-State connectivity.

Wide Area Network

A wide area network connects computers in different cities or countries. The network to connect computers that are thousands of miles apart is not built by an organization. Instead, the organization uses leased telephone lines.

It is obvious that WAN is not owned by a single organization. It is owned and managed collectively by many cooperating organizations.

Internet

Internet or Inter-Network is the connection of two or more networks so that a computer on one network can communicate with a computer on another network. The Internet is a set of thousands of networks linked together around the world. That is why it is also known as network of networks. The communication between two computers takes place such that the user does not have to worry about the technology used by the networks. It is easy to find out the location of another system on the network. The router (also called gateway) acts as an interface between two networks. The Internet has no central administration but there are protocols, which are followed by each network of Internet.

Intranet

An Intranet is a network that connects the internal computing resources of an organization. The prime motive of Intranet is to facilitate information sharing within the organization with the help of tools such as web browsers. For instance, a manager can check the inventory level sitting in his own room. Email becomes an easy way to communicate, without having to worry about the physical presence of the person on his seat. Intranets operate within the company's firewalls. A firewall is a method of isolating the company's computers behind a device that acts as a gatekeeper. All outgoing requests for information go to a special computer, which hides the sender's machine address but passes on the request. All incoming information is also checked

by the firewall computer. Employees can venture out into the Internet but unauthorized users cannot come in.

Activity B

What are some of the benefits of using a firewall for your LAN?

.....

.....

.....

.....

.....

.....

.....

4.5 ADVANTAGES OF USING INTRANET AND INTERNET

The present age is called the Information Age. In this information age, the strategic resource is information. Information is useful only when it is received in time. The communication networks have become the media of choice for transmitting the information. The time lag is so little that the data seem to travel in real time. Most of the business paradigms have changed due to communication networks. A team of doctors who are geographically apart can analyze the case of a patient simultaneously. The consumers are able to locate the item of their choice on the electronic shop, compare quality and prices before they place an order with a company. They expect the company to keep them informed about the status of their order. With little technology and proper management, the customer can be made happy. At the same time, the company can keep a record of their customers and send them information about their new schemes. The intranet facilitates online financial planning and project management. We can enumerate some advantages of using Internet and Intranet as follows.

1. With the help of Internet, an organization can create a web site where they post information about availability of products. The customers can log on to the site and check the availability without having to contact an individual in the organization. The company saves the staff cost and the customer saves time. The same mechanism can be used to interact with suppliers.
2. The intranet is used to send email, which is a cheaper way of contacting an individual in terms of time. The online financial planning and project management are additional facilities intranet can provide.
3. Collaboration project work by the team members who may be geographically separate is now a reality.
4. The time-to-market a product has reduced considerably. The delay in the product development phases has also reduced due to faster paperless communication.

5. Videoconferencing provides real-time interaction among people who are geographically apart. A company can locate its offices based on the availability of resources. The interaction among employees can take place through videoconferencing. The company will not have to spend extra on travel and at the same time, it can save money by locating its offices at optimum locations. Another advantage of a videoconference is that a person who may be somewhat shy or introvert can anonymously communicate his idea.
6. There are many dedicated discussion groups on the network. A person can join a group of his choice and share his knowledge with others and get help from others.
7. Entire world gets connected using the internet. All social media Apps, virtual classes in academic sessions, online shopping, eCommerce and unified communications are possible because of internet.
8. Browsing websites for study materials and research study, conducting online survey, Government to Citizens (G2C) applications, listening music online and watching videos are possible anytime from anywhere.

4.6 THE LAYERED ARCHITECTURE

Communication over a network is a complex task. The host that wants to transmit the data must know the address of the client, a route to communicate to the client. Then it must establish a connection and transmit the data. If an error occurs during the transmission, the client must detect it and inform the host so that it can re-transmit the data. It will make life simple if this complex task is divided into sub-tasks. There are two popular models- OSI and TCP/IP.

The International Standards Organization developed a framework for computer communication with an objective of providing a standard. The model provided by ISO is known as Open Systems Interconnection Reference Model, popularly as OSI model. The OSI model has seven layers. Each layer has a well-defined function and interface. Development and management of software has become simple due to the OSI model. The seven layers are briefly described here.

Application Layer

This is the top layer and provides user with an interface to the network. The primary purpose is to provide a set of utilities for application programs.

Presentation Layer

This layer formats the data for presentation to the user. This layer, for instance, might compress/uncompress the data.

Session Layer

This layer is responsible for initiating, maintaining, and terminating the network session on behalf of an application.

Transport Layer

This layer is responsible for creating data units called packets from the data that is to be transmitted. Each unit is appended with enough information

about its source and destination so that it can travel independently on the network without having to worry about other packets. The packets are combined at the destination into the original message. This layer makes sure that each packet has been received and without error. It may request re-transmission of a packet if required.

Network Layer

This layer accepts messages generated by the transport layer. Its responsibility is to route packets to the destination.

Data link Layer

This layer is responsible for initiating the physical transmission of the data. This layer also detects and corrects errors that might have occurred during transmission.

Physical Layer

This layer is concerned with transmission of data bits. This layer defines rules for transmission, such as voltage.

The TCP/IP model consists of four layers- application layers, network layer, data link layer, and physical layer. The network layer is equivalent to presentation layer, session layer, transport layer, and network layer of OSI model.

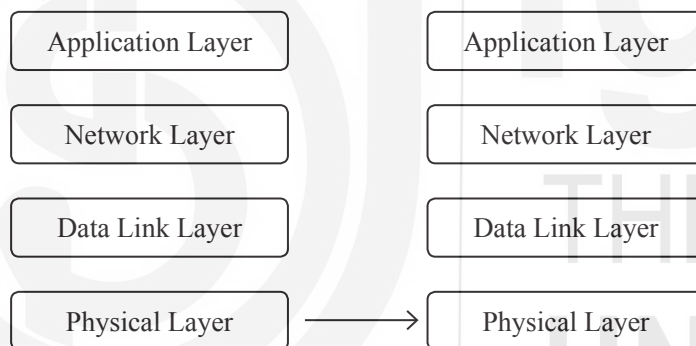


Figure 4.2: Four layers of TCP/IP model

We will briefly discuss Transmission Control Protocol/Internet Protocol next.

TCP/IP

TCP/IP is a protocol suite, which consists of protocols for all four layers. This suite is used for intranets as well as for Internet. 70% of all networks use TCP/IP. This protocol provides a reliable stream delivery and virtual connection service to the applications. The protocol ensures error-free transmission. TCP performs packetizing- large messages are broken up into smaller units called packets. Each packet contains the source and destination port identifier. A port is a logical entity that identifies an application. The packet also contains its sequence number and error checking information. The source and destination addresses are also included in the packet.

The address part needs some explanation. Each computer has some device such as network card through which computer is connected to the network. This device has an address. For instance, the Ethernet card addresses are 6 bytes long. The address is part of the hardware and cannot be changed.

This address is unique in the world. The network administrator assigns a 4 byte address to each computer on the network. This address is known as IP address. This address has to be unique on the network. The way this uniqueness is ensured is quite interesting. A computer on the LAN is assigned an address by the system administrator. These addresses are 4 numbers separated by a dot (.), each number is between 1 and 254.

The intranet addresses are usually 172.x.x.x. The system administrator maintains a table of assigned addresses and can easily ensure the uniqueness. A computer on the Internet is assigned an address by the Internet Network Information Center (InterNIC) who again maintains a table and can ensure uniqueness. The Internet service providers can get a block of addresses from InterNIC and assign them to individuals or organizations.

An address can be one of the following types:

Type A: The first byte is fixed by the NIC. The first bit of this byte is 0 for all type A addresses. The first byte can have value between 1 and 127. The remaining 24 bits are used to identify the host on the network.

An organization that has obtained a type A address may connect over 16 million computers to the Internet. Type A address is no more available.

Type B: The first two bytes are fixed by the NIC. The first two bits of the first byte are 10 for all type B addresses. The first byte can have value between 128 and 191. The remaining 16 bits are used to identify the host on the network.

An organization that has obtained a type B address may connect over 65000 computers to the Internet. Almost 80% type B addresses have already been allocated.

Type C: The first three bytes are fixed by the NIC. The first three bits of the first byte are 110 for all type C addresses. The first byte can have value between 192 and 233. The remaining 8 bits are used to identify the host on the network.

An organization that has obtained a type C address may connect 254 computers to the Internet. Almost 30% type C addresses have already been allocated.

There are Type D and Type E addresses also which are not for commercial use.

An organization has the freedom to divide its computers into various clusters and each cluster forms sub-network, referred to as subnet. The address of each computer on a subnet has same first three digits. The subnet mask identifies the subnet part of the address. The subnets are connected to each other by a gateway.

The IP address may be static or may be dynamic. A static address is assigned to the computers that are permanent members of the network. The computers that connect to the network through a modem are usually assigned IP address dynamically. Dynamic Host Configuration Protocol (DHCP) is part of the TCP/IP suite for dynamic addressing. A client can be assigned an IP address dynamically by DHCP server.

It is not so easy to remember 4 decimal numbers for each computer that one might like to communicate with. The name that a human being can remember easily is assigned to each server. These addresses are like `www.ignou.ac.in` or `ftp.iitk.ac.in` etc. These addresses are known as application layer addresses. A user wants to communicate to, let us say `ftp.iitk.ac.in` for a file transfer. This address must be converted to IP address and then to hardware address.

Domain Name Service protocol (DNS) translates server address to IP address. There are name servers, which maintain a special database of IP addresses and application layer addresses. There are thousands of name servers across the globe. If a computer does not know the IP address for an application layer address, it sends a request to name server. Other computers can reach a computer only if its address is a part of the database of at least one name server. This is the reason no one can just assign an address to its computer and expect to be found without the knowledge of InterNIC. Each computer can update its local database as more and more addresses from the name servers are learnt.

The **ARP** is used for converting an IP address to the hardware address. The protocol is based on broadcast; the computer trying to send a message to a particular IP address broadcasts a message asking for its hardware address. The computer with requested IP address responds back. The requesting computer also maintains a database of IP addresses and corresponding hardware addresses.

To send a message to a computer on another network, the sender needs to know a route to the destination. In static routing, the network administrator develops a routing table. In dynamic routing, the initial routing table is created by the system administrator as in case of static routing. The system then updates its routing table depending upon the situation of the network.

A message is broken up into packets. All the packets may travel independently on possibly different routes. This type of routing is known as connectionless routing. If all packets of a message are forced to travel same path, then a virtual circuit is set up. This type of routing is known as connection-oriented routing.

World Wide Web

The World Wide Web (WWW) made the Internet popular among people who were not part of the academic community. The Internet became popular primarily because of www applications. The prime objective of www applications is to facilitate sharing of text and image files. The www applications are implemented using client/server architecture. The server runs a software package called web server and clients run web browsers. The most popular web browsers are Netscape and Internet Explorer. Popular servers are Apache, Tomcat, Netscape server, and Microsoft server.

In order to get a page from the Web, the address of the desired page is typed into the browser. The user must know the address of the page in order to request the page. The web address is like any other address. It has components such as domain and computer name. Each computer on the Internet has a unique address. Each address is assigned by one of the address

assigning board such as InterNIC. These boards ensure that there are no duplicate addresses. Each address has computer name followed by domain name. Some of the well-known domains are as follows. The domain “edu” is reserved for educational institutes, “com” is for commercial organizations, “gov.in” is for government departments of India.

There are special software known as Search Engines which help a user in locating a web site. Popular Search Engines are Google, live.com, alexa.com, baidu.com, Bing/ MSN, Yahoo, Ask.com, duckduckgo.com, Ecosia. The user can run a search engine and type some keywords and the search engine locates many web sites, which have documents containing the keywords. The user may decide to then visit one or more sites. The sites listed by a search engine may run into millions and most of the sites may be irrelevant. Usually the top few sites are most relevant. The search engine maintains a database of all the web sites and this database is updated frequently.

The web server and the client both follow Hyper Text Transfer Protocol (HTTP) to communicate with each other. HTTP is an application level protocol for hypermedia documents. A virtual connection is established between the client and the server.

File Transfer Protocol (FTP) enables a host to send and receive files over the network- intranet as well as Internet. In a file transfer, there is a client that requests the transfer and there is a server, which entertains the request. There are many graphical FTP clients available these days, e.g. gftp on linux and ws-ftp on windows. A file transfer can be either closed or anonymous. In a closed file transfer, file transfer requires a valid login/password, which is given by the system administrator. An anonymous file transfer can be initiated by anyone. For a file transfer under ftp, a virtual connection between client and server is established to transmit control information. Another virtual connection is established for data transfer.

Telnet The telnet application enables a user on one computer to log on to another computer on the network. A virtual connection between the client and the server is established. The connection is used to transmit data. The protocol gives client an option to emulate graphics terminal.

Email Electronic mail is one of the oldest applications of Internet and it has become the most common way of communication. Email can be used to send a message to an individual or to a group of people. Text, formatted document, or an image in any format can be sent via email. The email uses client/server architecture. Clients are very advanced and provide amazing facilities. The user can create named folders and can save selected messages into the folder. It is almost like filing letters into a cabinet. A message can be sent to one person and a copy can be sent to another person. The message can be sent with a request for an acknowledgment. A message can be forwarded, deleted, or replied to. There are many protocols for email server as well as client. Simple Mail Transfer Protocol (SMTP) and Post Office Protocol (POP3) together form a complete mail transfer protocol. An SMTP server receives emails and stores them. The client can use POP3 to fetch and store emails in the individual local mailbox.

The SMTP server and POP3 client combination is quite popular these days.

A workstation running POP3 can dynamically access the SMTP mail server and retrieve the mails on it. A POP3 client makes a connection to the server either on demand or on regular intervals. If there is a mail, it is retrieved and the connection is aborted. POP3 seems inadequate if user would like to have the freedom to access his mailbox from anywhere. Internet Message Access Protocol (IMAP) allows a client to access and manipulate mailboxes on an SMTP server as local mailboxes. The mail stays on the server and user can access his mailbox from any computer running IMAP.

When you install a new computer on the network, you must obtain an IP address.

4.7 CLIENT/SERVER ARCHITECTURE

In the days of mainframes, the central server used to be a powerful machine. The software and the data both were stored on the main frame. There used to be a set of terminals that people used to connect to the server. This simple architecture is known as host-based architecture. Along the passage of time, the terminals were replaced by computers, which were not as powerful as the servers but were more powerful than the terminals. The server became the file and database server, which stored the data, but the application software ran on the client computers. As the applications became complicated, the network traffic increased and people started looking for an alternative. The client/server architecture was then proposed and it became very popular.

In the client/server model, a machine that provides a service is called a Server. The machine that makes use of the service is called a Client. For instance, server may store the database of an organization. A client connects to the database server for its needs. The client and the server work independently in a co-operative manner. The architecture facilitates the addition of a new client or server to the network. A server usually can handle many clients simultaneously. Therefore, most servers are powerful machines such as mainframes, minicomputers, workstations etc. The clients need not be as powerful as the servers and they usually are not. The clients and servers are on the network. But each server is assigned one or more specific tasks. For instance, a file server is used to manage user data files, a network server may control network connections. A database server stores the data of the organization and provides different levels of access to different clients. The servers provide their services through human readable names and well-known ports. When the client makes a request, it is not aware of the physical location of the server, or the processes of the server. The server details are hidden, but services are transparent. This is known as “service transparency”. It is an important feature of client/server architecture.

A request made by a client may involve data access, data processing, and result rendering to the client. The data access, the processing, and the presentation can all be done by the server or they can be shared between the server and client. The client/server implementation is called “distributed presentation” when all three components are on the server, and the client has part of presentation logic.

In remote presentation, the data storage and management and application logic is handled by the server, but the presentation logic is handled by the client.

In distributed logic arrangement, the data storage and management is handled by the server, the presentation is handled by the client and the application logic is shared between the server and the client.

In remote data management model, the data storage and management is the responsibility of the server and the other two functions are handled by the client.

In distributed database model of the client/server architecture, the client handles part of the data storage and management of application logic and presentation. The server complements data storage and management job of the client.

A client is called a thin client if it only handles the presentation job. A client is a fat client if it supports data storage, application processing, and presentation.

Client/server architecture is scalable. A new server can be added to the network, an application or part of it can be moved to another machine. The architecture supports heterogeneous hardware and software. One system may run MS Server and another one can run Linux and they can form a client/server pair. The failure of a network segment or a computer does not bring the entire system down. The client/server architecture reduces the unnecessary network traffic. The required part of the data stored on the server is transmitted to the client over the network. The file and data management is easier because only a single copy of data is stored at the server. The server controls the access and usually different people have different level of access.

4.8 CLOUD ARCHITECTURE

Cloud architecture has emerged as technology components that are combined to build a cloud. All the resources in cloud are pooled through virtualization technology and shared across a network. The components of cloud architecture include:

- A front-end platform (the client or device used to access the cloud)
- One or more back-end platforms (servers and storage)
- A cloud-based delivery methodology
- A network to connect cloud clients, servers, and storage

Together, these technologies create a cloud computing architecture on which applications can run, providing end-users with the ability to leverage the power of cloud resources. The concept of cloud computing is, however, very similar to mainframe computing, popular since the 1960s, where centralized servers ran applications that were accessed by 'dumb' terminals connected to a private network.

Cloud computing architecture enables organizations to reduce or eliminate their reliance on on-premises server, storage, and networking infrastructure.

Organizations adopting cloud architecture often shift IT resources to the public cloud, eliminating the need for on-premises servers and storage. This reduces overhead cost and the need for IT data center, cooling, and power requirements. Instead, the overhead cost is replaced them with a monthly

IT expenditure. With mutual agreement the payment terms may be even quarterly, half-yearly or annually.

This shift from capital expenditure to operating expense is a major reason for the popularity of cloud computing today. There are three major models of cloud architecture that are driving organizations to the cloud. Each of these has its own benefits and key features.

- **Software as a Service (SaaS):** SaaS architecture providers deliver and maintain applications and software over the Internet. Thus, it eliminating the need for deploying the software locally. SaaS applications are typically accessed via a web interface. Software solutions in terms of ERP, SAP, CRM, SCM, PMS and others ,say, related to hospital and patient management, virtual meetings and conducting training programmes on line are possible at lower cost.
- **Platform as a Service (PaaS):** In this cloud model, the service provider offers a computing platform and solution stack, often including middleware, as a service. Organizations can build upon that platform to create an application or service. The cloud service provider delivers the networks, servers and storage required to host an application while the end user oversees software deployment and configuration settings.
- **Infrastructure as a Service (IaaS):** This is the simplest form of cloud architecture. In this form, a third-party provider eliminates the need for organizations to purchase servers, networks or storage devices by providing the necessary infrastructure. In turn, organizations manage their software and applications, and only pay for the capacity in terms of servers and storage they need at any given time towards hardware.

There are many IT service providers to support cloud architecture and provide the facility to organizations. One such example is VMware. It provides unified approach to build, run, and manage traditional and modern applications on any cloud. As a single platform, VMware functions across all applications and multiple cloud environments helping organizations to migrate and run applications seamlessly.

- VMware allows enterprises to create application modernization and multi-cloud strategies that support cloud operations across a multi-cloud landscape.
- VMware offers a broad set of capabilities to build, run, and manage modern apps on any cloud through a rich platform as a service (PaaS) layer.
- VMware also enables better management of multi-cloud architecture with comprehensive visibility across Amazon Web Services (AWS), Microsoft Azure, Google, and Oracle public clouds; Kubernetes; and VMware-based private and hybrid clouds.

4.9 SUMMARY

This unit has given you an introduction to the data communication networks and their importance to an organization. The components of a network are described which are same for all types of networks. Various types of network

were introduced. A network operating system is also an integral part of a network. The protocols- OSI model and TCP/IP are the main protocols, which were described in detail. The TCP/IP is a suite of protocols which has application layer protocols, network layer protocols, data link layer protocols and physical layer protocols. The application layer protocols and network layer protocols were described in some detail. The client server architecture and cloud architecture were also explained.

4.10 SELF-ASSESSMENT EXERCISES

1. Describe the main components of a network including hardware and software.
2. Describe various types of networks.
3. Distinguish between LAN, and MAN.
4. Distinguish between SWAN and WAN.
5. Distinguish between Internet and Intranet.
6. Write a detailed note on the advantages of networks.
7. Describe TCP/IP protocol suite briefly. You may include any five protocols in your note.
8. Describe Client and Server architecture.
9. Describe Cloud architecture.
10. Distinguish between Client-Server and Cloud architectures.

4.11 REFERENCES AND SUGGESTED FURTHER READINGS

FitzGerald, J. and Dennis, A., *Business Data Communications and Networking*, John Wiley & Sons, Inc.

Tanenbaum, A.S., *Computer Networks*, Prentice Hall PTR.

<https://www.section.io/engineering-education>, site accessed on November 18, 2021

<https://www.vmware.com/topics/glossary>, site accessed on November 19, 2021

Block

2

INFORMATION SYSTEMS

UNIT 5

MIS and Control Systems

65

UNIT 6

Information systems Economics & Security

82

UNIT 7

Transaction Processing Systems, DSS, and EIS

98

UNIT 8

Integrated Applications

129

BLOCK 2: INFORMATION SYSTEMS

Unit 5: MIS and Control Systems enables you to comprehend the evolution of MIS and understand the main features of a Management Information System (MIS). You will be able to have an idea about the development of MIS and its role at various management levels in an organization. The unit illustrates Anthony and Simon framework for understanding the MIS and decision-making process. You will be competent enough to appreciate different viewpoints regarding applications of MIS in organizations. Also you will understand about system vulnerability, control mechanism and requirement of auditing.

Unit 6: Information systems Economics & Security begins with valuing the growing need for information. It reveals the relationship between data and information, information and knowledge. You will be familiar with the concepts like cost and value of information, ethics in information security, right to information and how to protect the information systems. The unit finally portrays the main reasons for success and failure of an MIS.

Unit 7: Transaction Processing Systems, DSS, and EIS facilitates you to comprehend and conceptualize about transaction processing systems (TPS), decision support systems, executive information systems with various illustrations about TPS with required inputs processing and desired outputs at various levels from the respective systems. You will understand about the decision making process. This Unit also brings to you the differences between MIS, DSS, and EIS..

Unit 8: Integrated Applications entails discussion on importance of integrated systems, their increasing use in organizations with advantages of using them. Illustrations have been provided here in this Unit of some business solutions of integrated systems by big players like Microsoft, SAP, Oracle. Business utility of integrated systems have been discussed too.

UNIT 5: MIS AND CONTROL SYSTEMS

Structure

- 5.1 Management Information Systems – An Introduction
- 5.2 Objectives
- 5.3 System View and Status of MIS in Organizations
- 5.4 Framework for Understanding MIS and its Role at Various Management Levels
- 5.5 Organization and Information System-Two way relationship
- 5.6 MIS as Decision Assisting Tool
- 5.7 System Vulnerability, Abuse, and Auditing
- 5.8 Summary
- 5.9 Unit End Exercises
- 5.10 References and Suggested Further Readings

5.1 MANAGEMENT INFORMATION SYSTEM – AN INTRODUCTION

The subject of Management Information System (MIS) has different meaning for different people. MIS, initially, were built to process transactional data of an organization and to produce regular reports. With little more advancement the system generated a report in a suitable format that created an impact on its user and provoked an action, a decision or an investigation. Today, an information system has evolved to the stage where they handle databases and secondary data available on the web to facilitate decision-making. Accordingly, definition of MIS has also evolved. There are many closely related definitions in use. The terms MIS is synonymously used with terms the Information System (IS), the Information and Decision System and the Computer based Information System.

The MIS is defined as an integrated system of man and machine for providing the information to support the operations, the management, and the decision-making function in the organization.

The above definition emphasizes an association between MIS and decision-making. An application software that processes data, which is not used for decision-making, cannot be called an MIS. For instance, a computer-aided design system is not an MIS.

An MIS deals with information that is systematically and routinely collected in accordance with a well-defined set of rules. In other words, data collection is a planned activity for which resources are allocated and rules are defined.

The information provided by an MIS assists managers in planning, organizing, staffing, coordinating, directing and controlling the operations of an organization. The management experts have viewed these steps as

Management Control system. Figure 5.1 shows the relationship between operations, planning and control.

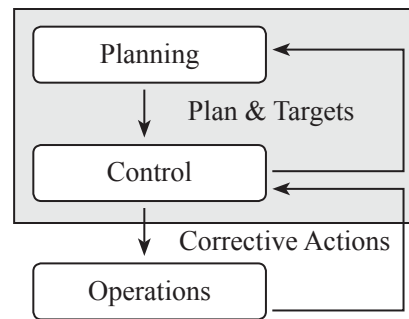


Figure 5.1: Management Control System

In any organization that has planned activities leading to the achievement of the stated goals, there is always a control process in place that measures progress towards these goals and enables the manager to deduct the deviations from the original plan in time. It is the responsibility of the management to take corrective actions before it is too late. The deviations may be due to environmental changes or due to the mistakes made by people. An MIS is concerned with planning and control. An MIS has large amount of data as its integral part that is stored and managed by a data base management system.

The exponential growth of information all around makes it necessary that information is probably collected, stored and retrieved in various fields so that it could be usefully exploited whether and when needed. The concern is “What information does the manager need to manage effectively”? We are interested in a system for providing the necessary management information. MIS is a system that aids management in making, and implementing decision. An MIS must have the following features:

1. It must be capable of handling voluminous data. The data as well as transactions must be validated.
2. It must be able to perform operations on the data irrespective of the complexity of the operations. Often time multi-dimensional analysis is required.
3. An MIS should facilitate quick search and retrieval of information. An MIS must support mass storage of data and information.
4. The information must be communicated to the recipient in time. Moreover, the communicated information must be relevant.
5. The information system must always have internet link to dig out related information and data from the web for better analysis helping the management in taking decision.

5.2 OBJECTIVES

After reading this unit, you should be able to:

- Identify the main features of a Management Information System (MIS);

- Explain the Anthony and Simon framework for understanding the MIS and decision-making process;
- Explain the basic overview of the systems concept;
- Describe the management functions at various levels in the context of relationships between management and informational needs;
- Use MIS as a technique for making programmed decision;
- Explain the vulnerability of an information system; and
- Discuss the control and audit structure required for an information system.

5.3 SYSTEM VIEW AND STATUS OF MIS IN ORGANIZATIONS

Today besides professional managers, all administrators, politicians, academicians have become aware of the need for adopting an integrated holistic perspective by using the systems approach to problem-conceptualization and decision-implementation. Today we find everyone talking of systems - the transport system, educational system, healthcare delivery system, defence system, economic system, communication system, management information system, transaction processing system, decision support system, computer systems, etc. But what exactly do we mean by a system? Let us elaborate to know about.

What is a system?

It is an entity; conceptual or physical, which consists of interdependent parts or components. It is this interdependency which is characteristic of the parts of the system. A system is a complex of elements or components directly or indirectly related in a casual network. This brings in the notion of some type of feedback and control to see whether or not the system is a position to achieve the goals/purpose/objectives of the system. Any system must have an objective or a set of objectives or a hierarchical set of objectives. In a large context, a system is an assembly of procedures, processes, methods, routines techniques etc. united by some form of regulated interaction to form an organized whole. In fact no system, unless it be a totally closed system, can exist in isolation.

A system is made up of sub-systems, which may be composed of further sub-systems. We could carry on this refinement till we arrive at the so-called 'black box' level, which is some perceptible manageable level. Just as system is made up of sub or sub-sub-system, it itself is part of a super or supra system. This could be termed as the environment in which the system operates.

We can graphically depict the above narrative description in the form of Figure 5.2 below. Let us give an illustration in the context of Figure 5.2. One could think of an industrial system or a factory system. A factory system has various sub-systems like the production sub-system, the shipping sub-system, the financial sub-system, the marketing sub-system and the

personnel sub-system. Now a production sub-system could consist of sub-sub-systems of production control, materials control, quality control etc. the materials sub-sub-system can be further broken down into ‘black boxes’ say purchasing, stores, transportation etc. In turn, the factory system is part of the larger economic system of the country which would be the so-called superior or supra system.

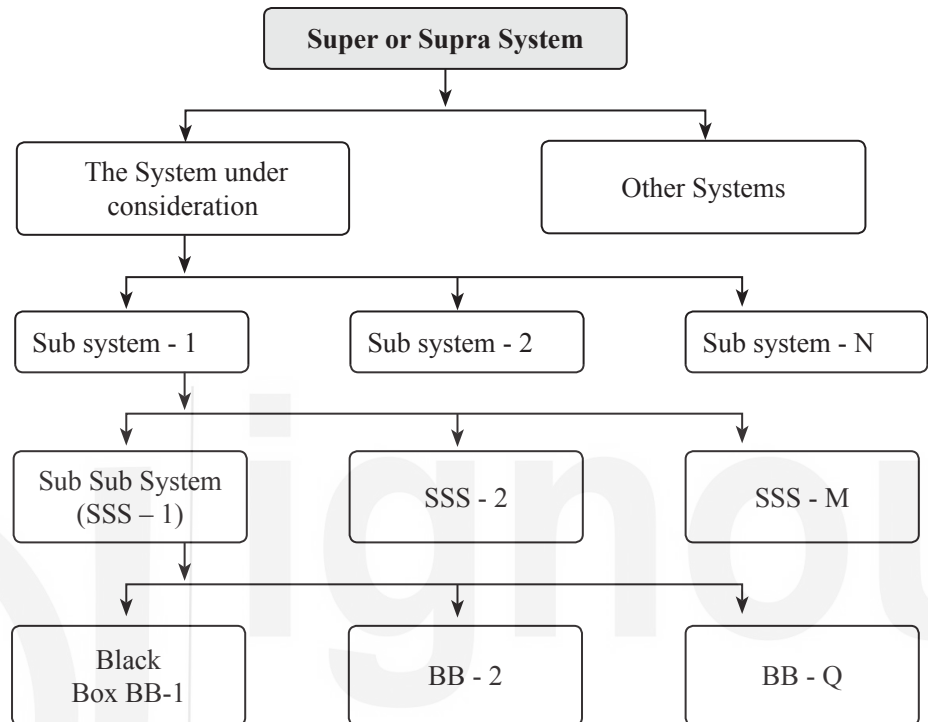


Figure 5.2: Hierarchy of Systems

Activity A

Think of at least three examples in the context of the Figure 5.2

.....

.....

.....

.....

Let us have another example in the context of Figure 5.3 below. Data could be a set of input into a data processing, which would process/transform/convert the data into output or information. One could think of raw material entering as input into production system, which is converted/transformed or processed into an output i.e. some final product.

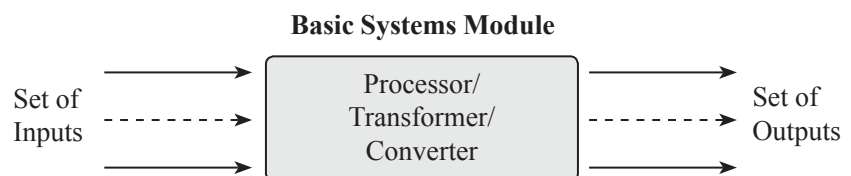


Figure 5.3 (a)

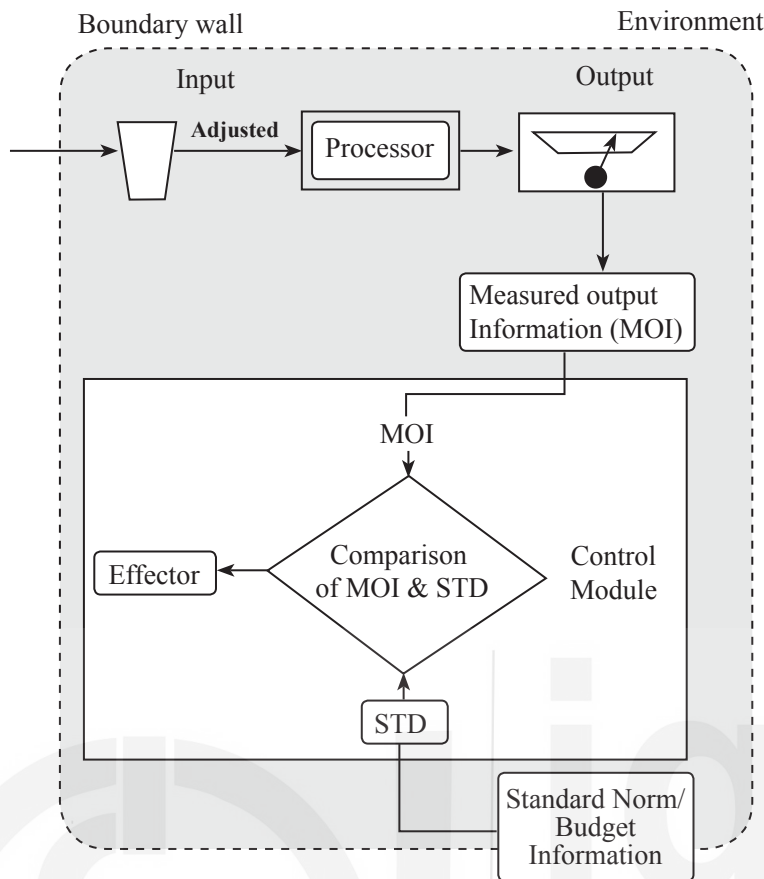


Figure 5.3 (b)

Activity B

In the context of Figure 5.3, give at least three examples mentioning the inputs, the processor, and the set outputs.

Input	Processor	Output
.....		
.....		
.....		

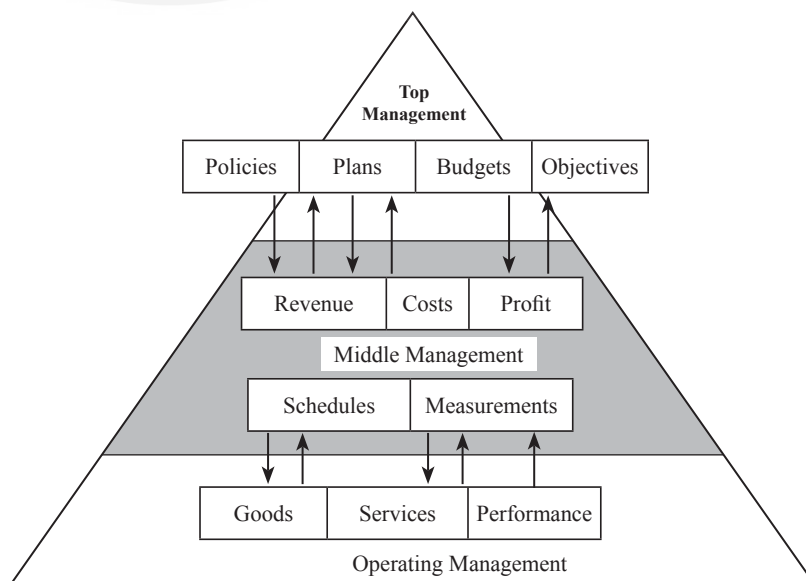


Figure 5.4: Interaction of Management Levels

Continuing the example further in the context of Figure 5.4 this time, the quality of the finished product could be measured by comparing it with the standard specifications of the finished product. Depending on the deviations or variances the manager can then adjust the quality and quantity of the raw materials. The environment of the factory system under consideration could be other factories, competitors, customers, markets, sociopolitical and cultural factors, government etc.

Information systems are used in all functional areas and operating divisions of business. In finance and accounting, information systems are used to forecast revenue and business activity, determine the best sources and uses of funds. Information systems have been used for managing cash and other financial resources, and analyzing investment. Financial health of an organization is also checked using IS. In sales and marketing, information systems are used to develop new goods and services (product analysis), determining the best location for production and distribution facilities (site analysis), determine the best advertising and sales approaches (promotion analysis) and set product prices to get the highest total revenues (price analysis).

In manufacturing, information systems are used to process customer orders, develop production schedules; control inventory levels and monitor product quality. Service industries such as airline industry and railways use information systems to serve their customers better. Banks and other investment firms' use IS to make good investments and sanction sound loans. Publishing houses, healthcare organizations, and retail companies all make use of information systems to serve their customers better and maximize their profit.

Activity C

Detail some functional MIS systems in your organization or any organization of your choice. What are their distinguishing characteristics? Why are they failures or successes?

.....

.....

.....

5.4 FRAMEWORK FOR UNDERSTANDING MIS AND ITS ROLE AT VARIOUS MANAGEMENT LEVELS

There is too much data and information in an organization. In order to design a MIS successfully, we need a framework to structure the information so that the data and information relevant for decision-making can be separated from rest of the data. Before we talk about the design of MIS, let us understand the strategic management of a business. An organization must respond to market forces, competition, to environment and to technological changes. The scope of business is wide, touching many fronts. A business, among other activities, must do a long-term strategic planning. There are many methodologies for strategic planning. According to model presented

by **Robert Anthony**, the strategic planning is one of the major activities in business planning and control. The other two are the management control and operational control. This framework is illustrated in Figure 5.5.

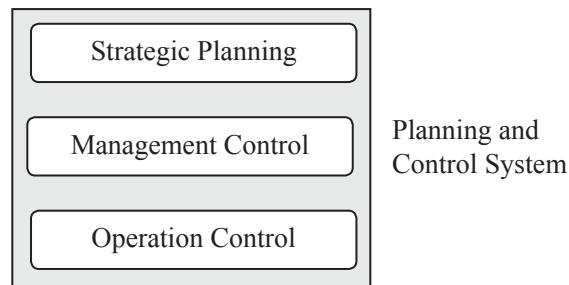


Figure 5.5: Framework of business planning and control

1. Strategic Planning is the process of deciding objectives of the organization, determining the possible shift in objectives, deciding on the resources used to attain their objectives and the policies that govern the acquisition, use and disposition of their resources.
2. Management Control is the process by which managers assure that the resources are obtained and used effectively and efficiently to attain the objectives of the organization.
3. Operational Control is the process of assuming that specific tasks are carried out effectively and efficiently.

It is useful to classify the above definitions with some examples. The table below gives instances of planning and control activities in different functional areas.

Table 5.1: Planning and Control Activities in Different Functional Areas

Functional Areas	← ACTIVITIES →		
	Strategic Planning	Management Control	Operational Control
Production	Location of a new factory	Determine the product mix for a monthly production program	Scheduling specific jobs on specific machines in a shift
Marketing	Entering the export market	Media planning for advertising expenditure	Planning sales contacts to be made by a salesman in the next week
Finance	Raising capital by issuing new shares	Determining maximum levels of credit for customers	Determining what action to take against nonpayment by a specific customer
Personnel/HR	Deciding on changes to be made in the organization structure	Determining who will be promoted to fill a vacated post at middle and lower levels, in the organization.	Determining which workers will be on each shift.

Anthony's framework enables us to understand the characteristics of information needed to support the three types of planning and control process. The table 5.2 below depicts these characteristics and highlights

the substantial differences in information required for strategic planning, management control, and operational control.

Table 5.2: Differences in Information required for three types of Planning and Control Processes

S. No.	Information Characteristic	Strategic Planning	Management Control	Operational Control
1	<i>Volume</i>	Low	Intermediate	High
2	<i>Level of Aggregation</i>	High	Intermediate	Low
3	<i>Frequency of use of a particular type of data</i>	Low	Intermediate	High
4	<i>Currency requirement</i>	Low	Intermediate	High
5	<i>Accuracy</i>	Low	Intermediate	High
6	<i>Scope</i>	Wide	Intermediate	High
7	<i>Source</i>	Significant amount from external sources	Mostly Internal	Entirely Internal
8	<i>Predictability with user</i>	Low	Fairly High	Very High
9	<i>Variability with user</i>	High	Intermediate	Low
10	<i>Distance of user (in Organizational terms) from sources within organization</i>	Fair	Fairly close	Close

Let us now look at **Simon's framework** that has broken down the process of decision making into three stages:

1. *Intelligence*: This is the stage in which the decision maker recognizes that there is a problem or opportunity that requires him to make a decision.
2. *Design*: The decision maker determines the alternatives that are available to him to resolve the problem or exploit the opportunity.
3. *Choice*: In this stage, an alternative generated in stage-2 is singled out to be pursued. The selection process may involve feasibility analysis or cost-benefit analysis.

With this framework, we can distinguish between three major classes of decisions.

- a) **Programmed Decisions** are there in which all stages are handled by following a preset well-defined procedure. The decisions are repetitive and routine which arise often and are capable of being modeled mathematically in their entirety. The classic example would be inventory-ordering decisions.
- b) **Non-programmed decisions** are difficult to structure in logical-mathematical terms. These decisions cannot be handled in well-defined and pre-specified procedures. These opportunities are not repetitive in nature and they require fresh intelligence, design and choice phases to be executed. An example would be the decision to set up a new factory or launch a new line of product.
- c) **Semi-programmed decisions** are those in which at least one and no more than two of the above stages can be handled by well-defined preset procedures. An example where the intelligence phase is

well structured would be the diverse kinds of variance analysis. A comparison with a budget or standard is undertaken in a well-defined way to signal the need for a decision. Subsequent stages of design and choice, however, are not handled by a set procedure.

Recall that an MIS supports problem-specific decision-making. Depending on the framework used by the organization for decision-making and goal set for MIS, the designer should determine the information needs.

An MIS should not automate the existing procedures. MIS should act as a catalyst of change in the processes of an organization. For instance, a private bank sanctions loans by using a sequential process. An applicant applies for a loan, the details provided by him are verified, and his application details are entered into the bank application format along with his credit limit. The computer application is then passed on to the loan sanctioning authority. The process takes two weeks time even though the staff spends about thirty minutes on the application. A workflow system should not automate the existing process. It should aim to reduce the application processing time to less than a week. The present workflow systems let everyone look at the application simultaneously and each concerned person adds his feedback. The sequential process has been changed to a parallel process. The total time has come down to less than a week.

According to **Zani**, the important determinants of MIS design are:

1. Opportunities and risks
2. Company strategy
3. Company structure
4. Management and decision-making process
5. Available technology
6. Available information sources.

An MIS should be designed, viewing the organization. A company's structure sub-divides essential tasks to be performed, assigns them to individuals, and spells out the interrelationships of their tasks. The organizational structure and the tasks determine the information needs of the company.

The MIS designer must plan to deliver reports in line with the organization structure. This means that the main decision makers and the power centers must be recognized in the MIS. If the decision-making responsibilities are clearly defined and allocated in the organization, MIS must capture them. If the organization culture provides sufficient incentives for efficiency and results, the MIS support this culture by providing such information, which will aid the promotion of efficiency.

The organization system is an open system and MIS should be so designed that it highlights the changes to the concerned level in the organization so that the action can be taken to correct the situation.

The designer of the MIS should take care of the data problems. The input data to the MIS may contain bias and error. The inputs to the MIS must be controlled to ensure impartiality, reliability and consistency.

If the organization culture provides sufficient incentives for efficiency and results, the MIS should provide information that will aid the promotion of efficiency.

If the organization is an open system then MIS should be designed to highlight critical changes in the system or in its environment.

In designing an MIS there are two types of situations one may come across. If the organization has no experience of computing applications, which will create the maximum impact on the organization, it can be identified by using Zani's framework. Key success variables are however seldom obtained through a questionnaire survey of managers. Data on environment, past company performance must be analyzed and discussed to identify key success variable. It is sometimes useful to pen down a quantitative measure of such variable. For example the performance of a textile unit can be summed up through two indicators: contribution per loom shift and fixed cost per loom shift. Similarly the performance of a shipping company may be measured as gross operating profit per day per voyage. Precise definitions of performance indicators enable the analyst to understand and quantify the likely impact of improvement in different task of planning and monitoring.

An analysis of the company's key success variables can be done only after a thorough understanding of the company's operations. Consultants and vendors who do not spend adequate time in understanding the operations are unlikely to throw up application areas, which will create the maximum impact. They are likely to suggest "off-the-shelf" applications. One should use standard software, which is available for such applications.

For a company getting into automation using information system, a list of applications would have to be generated, keeping in view the future perspective, a least for five years, on the basis of which a suitable configuration would be decided. However the development and implementation of the applications would have to be done in a phased manner. The first few applications must be those, which can create an impact on the performance of the organization, are quick to implement with the least amount of changes in the existing procedures and systems.

By and large an effort is made to create useful databases, which capture data during the execution of routine data processing systems. Such data are then analyzed to produce periodic planning report for monitoring.

Examples of such systems are the sales analysis based on invoice processing; inventory control based on stock accounting; costing and profitability analysis on the basis of financial accounting system. Marginal additions to data fields, new coding structure, and revised procedures are introduced to make the data base and reporting more useful.

Factors Facilitating Implementation of MIS

A few factors, which will increase the chances of a successful implementation of MIS, are:

- 1) Involvement of top management in the computerization effort, in defining the purpose and goals of computers within the organization.
- 2) Selection of an IT Manager who is well versed with IT tools and also has the skills to involve managers in choosing application areas, identifying information needs and designing reports.

- 3) A computer staff, which has interdisciplinary skills in computers, management, and operations research.
- 4) A balanced expenditure on hardware and software.

Now let us look into the role of MIS at various management levels. We come across saying that management can be understood by observing what managers do. Management can also be understood by the type of functions a manager performs. A manager usually performs the following functions: planning, organizing, Staffing, Directing, Coordination, Feedback, Reporting and Budgeting. In fact management is a process of achieving an organization's goal and objectives by judiciously making use of resources of men, material, machines, money, methods, messages and moments (the last two in the context of information being vital resources to the manager/decision-maker).

Management can also be seen as structured into three hierarchical levels namely, top level, middle level and bottom level or strategic, tactical and operational levels, respectively. Although lines of demarcation are not absolute and clear-cut, one can usually distinguish certain layers within the organization, which are characterized by, the classical pyramidal type of structures as shown in Figure 5.4. Top management establishes the policies, plans and objectives of the company as well as a budget framework under which the various departments will operate. These factors are promulgated and passed down to middle management. They are translated into cost or profit centre concept. These are reviewed, analyzed and modified in accordance with the overall plans and policies until agreement is reached. Middle management then issues the specific schedules and measurement yardsticks to the operational management. The operational levels has the responsibility of producing goods and services to meet the revenue, profit and other goals, which in turn will enable the organization achieves its overall and objectives.

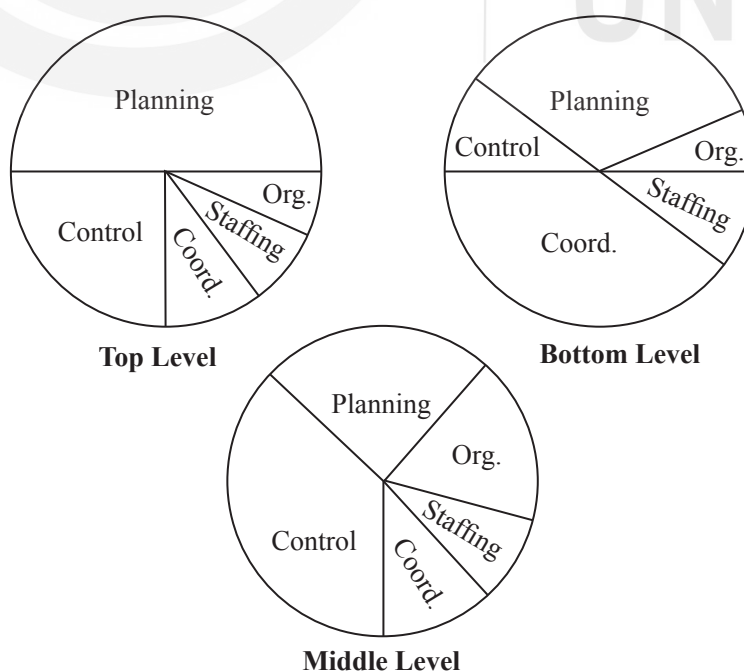


Figure 5.6: The Allocation of Managers' Time

The hierarchical view of management is important for two reasons: information needs tend to be different at different levels of management and the amount of time devoted to any given function varies considerably with the level as can be seen in Figure 5.6. The job content at various management levels is further elaborated in Table 5.3.

In the context of MIS, management can perhaps be best defined as a process of (i) selection of objectives (ii) judicious allocation of resources (iii) determining operational plans and schedules (iv) keeping control of progress and (v) evaluation through feedback. Each of these areas requires certain decisions to be made.

Thus we take strategic decisions at the top level, tactical decisions at the middle and operational decisions at the junior level. As can be seen from Table 5.3, the type of problems and decisions at the junior level are quite deterministic and structured, so we can have programmed decisions.

Table 5.3: Job Content of Management Levels

S. No.	Character	Top Management	Middle Management	Operating Management
1	<i>Focus on Planning</i>	Heavy	Moderate	Minimum
2	<i>Focus on Control</i>	Moderate	Heavy	Heavy
3	<i>Time Frame</i>	1-5 years	Up to 1 year	Day to Day
4	<i>Scope of Activity</i>	Broad	Entire functional area	Single sub-function
5	<i>Nature of Activity</i>	Relatively	Moderately	Highly
6	<i>Level of Complexity</i>	Very Complex, many variables	Less complex, better defined variables	Straightforward
7	<i>Job Measurement</i>	Difficult	Less Difficult	Relatively Easy
8	<i>Result of activity</i>	Plan, policies & strategies	Implementation, schedules, performance yardsticks	End-product
9	<i>Type of Information utilized</i>	External	Internal, reasonable accuracy	Internal, historical level of accuracy
10	<i>Mental Attributes</i>	Creative innovative	Responsible, persuasive, administrative	Efficient, effective
11	<i>Number of People Involved</i>	Few	Moderate number	Many
12	<i>Department/ Divisional interaction</i>	Intra-division	Intra-division, Inter dependent	Intra-department

Source: J.Kanters-"Management Information Systems", Prentice Hall: Englewood Cliffs

Through the classic pyramidal structure is generally acceptable; unfortunately in the modern complex organization this neat, militaristic,

configuration seldom (!) fits the reality. The modern manager must be capable of managing his/her information systems for strategic planning, management control, and operational control.

Activity D

Can you visualize MIS without computers? Justify your answer.

.....

.....

.....

.....

.....

5.5 ORGANIZATION AND INFORMATION SYSTEM – TWO-WAY RELATIONSHIP

A system is an assembly of elements arranged in a logical order to active certain objectives. An organization is also a system. H. J. Leavitt advocates that an organization should be viewed as a socio-technical system consisting of people, tasks, technology, culture and structure. The modified Leavitt's model is shown in the Figure 5.7 below:

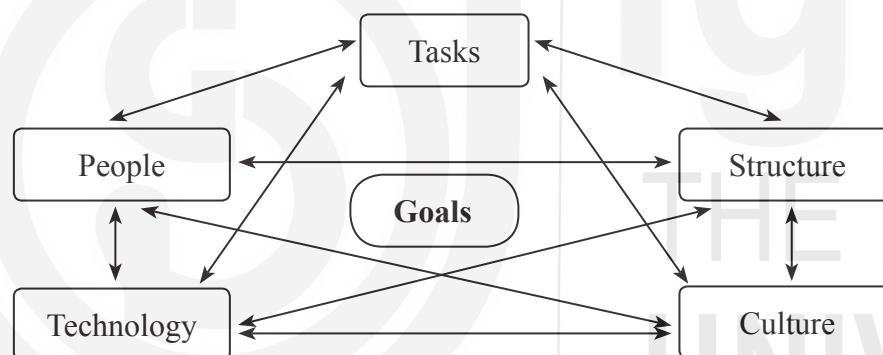


Figure 5.7: An organization as a socio-technical system

An organization is an open system that has the capacity to adjust itself to the changing environment. The goals of an organization change in response to the changes in organization or in its environment. The organization must change as system to stay in tune with the goals.

MIS should be designed viewing the organization as a system. MIS design should give due weightage to the human side of the organization and its culture. MIS should be designed to give reports to main decision maker. In other words, the designer must study the organization structure and identify the power centers.

In a tall hierarchy with a high degree of centralization, the MIS should give control information to the higher management.

If the organization is structured on a functional basis, then the MIS should have a functional design.

If the organization works on a standardized system where rules, policies, systems and procedures have been laid down, then there become part of the MIS.

5.6 MIS AS DECISION ASSISTING TOOL

We have studied above that Simon's Model divides decision-making into three phases - Intelligence, Design and Choice. In the intelligence phase, potential problems and opportunities are identified. In the design phase, alternative solution to the problem is developed. In the choice stage, a course of action is reflected.

In certain cases, the decision can be made using a rule, procedure or quantitative method. Such problems are known as structured problem. For example, an organization may decide to place a purchase order for every purchase requisition without worrying about merging them. Such decision can be easily programmed. The organization may have a fixed re-order point and fixed re-order quantity irrespective of demand. Automated system such as transaction processing systems and MIS are often used to handle programmed decisions. Such systems generate reports for concerned people so that they can take action. The decision taken is known as 'Programmed Decision'.

If the problem is unstructured, the solution cannot be arrived at using a set of rules or procedures. The rules and procedures, at best, can help in identifying alternatives. An information system that assists decision maker in making a non-programmed decision is referred to as Decision Support System (DSS). A decision support system usually has large amount of data that is managed by a Database Management System. The DSS presents various views of data to the decision maker to facilitate decision-making.

The views and reports may present details according to the needs of the user. A user may prefer graph whereas another user may like to see data in tabular form. A DSS is designed to provide presentation flexibility to its users. Details can be rolled up or drilled down depending on the requirement. For instance, sales manager of area A would be interested in knowing total sales, details of each sub-area within the main area. The manager would also like to know sales in area B, area C etc. to measure the relative performance. The manager would not be interested in details of sub-areas of area B as C; and will look down area A to get complete details.

DSS should provide facility to its user to perform statistical analysis of data. The sales data, for instance, can be analyzed for identifying seasonal fluctuations from regular demand change. It should also have provision to dig out secondary data from web. There are whole lots of software packages that can do amazingly good analysis of the data. Let us say, a company is considering an investment in an IT Project. It can do a what-if-analysis to analyze worst and best scenario. The expected cost and benefit can be changed within limits to estimate pay back period.

These are situations when a closed form solution is not available. In such situation a simulation model is constructed to gain insight. The model is tested on large number of inputs. Sometimes, simulation is done to study the behavior of system over a period of months in few days. A simulation system assists manager decision-making.

A decision support system may help a manager in perform goal-seeking analysis. In goal seeking analysis, you work backward starting with goal

to arrive at conditions required to achieve that goal. For instance, you may allocate budget for inventory and then backward to figure out the stock level, re-order point, delivery time etc.

Ideally, one would like to have a DSS to help in decision making in every situation. So far, no DSS has come close to being so general, perfect, and useful at the same time. A DSS, in practice, is developed for assisting in decision-making process for specific problems.

5.7 SYSTEM VULNERABILITY, ABUSE, AND AUDITING

Due to the potential of information systems, they are under constant attack from intruders and hackers. Unauthorized access to data and programs may cost an organization a fortune. For instance, an individual may disclose the stock levels of an organization to a potential buyer who may decide to go to the competitor if he feels that you may not be able to fulfill the order in time. In the world of e-Commerce, and internet easily accessible, one mistake could cost a client enough to justify a lawsuit. To avoid, abuse of information system, proper controls must be set. There are many threats to information systems such as hardware failure, software failure, user errors, software changes, theft of data, services, equipment, and telecommunications problems.

Systems become vulnerable because of system complexity. Any disaster has an extensive effect. System vulnerabilities differ from person to person.

- User: Identification, Authentication, and Subtle Software Modification
- Programmer: Disables Protective Features; Reveals Protective Measures
- Maintenance Staff: Disables Hardware Devices; Uses Stand-alone Utilities
- System Operator: Doesn't Notify Supervisor, Reveals Protective Measures, disables or changes the network setup etc.

It is better to think of all possible misuse and fraud from within the organization and establish system control to prevent them. Such controls are reformed to as deterrence controls. We next discuss some of these controls.

Entry of people who use the system should be monitored through biometrics systems such as finger print identifier, retina scanner, voice entry. Smart card identifiers and sanitization system also may be used.

The standard procedure such as well chosen passwords and different levels of access are always put in place.

Any output generated should leave an audit trail to identify any misuse of the reports. All transactions through the information system should have proper audit trail.

The database management system provides mechanism to give different level of access to different users to protect data. With proper configuration and use of encryption, the network misuse can be prevented.

An information system is like any other system in the organization. It must be audited and reviewed to make sure that it is operating and being used

as intended. The initial requirements and objective document serves as a reference. The system is audited for its performance, use, cost, benefit and other design criteria such as planned activity for which resources are allocated and time frame is set. A planned review is time-driven that may be done every 6 months or every year. The review must have a clear objective and reviewers must produce a written report. The report must comment on the following factors:

Objective: Is in tune with the organizational goal?

Hardware/Software/Network/DBMS: Are they being used as intended? Will they be able to meet requirement for next few years? Is there any bottleneck?

Staff: Is staff sufficient and sufficiently trained? Are People controllable using the system? In case of any skill gap because of version changes in systems, proper training must be provisioned to bring the staff at the current level of the updated system.

Safety: Are enough controls present to ensure safety? Are the controls working effectively?

Cost Benefit: This is one of the key reasons for taking up review activity. This is also the most difficult part of the review process. The reviews should try to be as objective as possible.

5.8 SUMMARY

This unit has given you a fair understanding of the main feature of a Management Information System in Organizational Context, describing its various functions, importance and relationship with planning, control and operations in an organization i.e. what an MIS is and what it is not.

Further, the unit systematically leads you to the prevailing status of MIS in organizations, discussing the ever growing need of information and its proper handling (processing), which in turn led to the development of MIS and advent of computers therein to cope with the hazards faced in coordinating and managing the organizational challenges. We have also discussed different viewpoints about the MIS given by same management scientists.

In this unit, we have learnt the basic structure of an information system. We also learnt the MIS as an organizational unit within an organization. MIS and its role at various levels of management were also discussed. MIS is used for making programmed decision as well as in assisting the decision making process. The information system is vulnerable and it can easily be abused. An information system requires a control structure and it needs regular audits.

5.9 UNIT END EXERCISES

1. Define MIS. What are the main features of an MIS?
2. What are the functions that MIS supports in an organization?
3. What is Simon's framework for decision-making? How does it help in MIS design?
4. What are the determinants of MIS design according to Zani?

5. Explain the following statement, “there is a two way relationship between organization and Information Systems”.
6. Discuss fully the purpose of an information system.
7. “In the end, the information system is recognized only as a foundation for human judgment, insight and inventiveness”. Discuss.
8. What impact does the implementation of programmed decision-making have on the management system of an organization?
9. Compare and contrast programmed decision making information system and decision assisting information system.
10. What are different types of controls and audits required for an information system?

5.10 REFERENCES AND SUGGESTED FURTHER READINGS

Davis, G.B., *MIS Conceptual Foundations, Structure and Development*, McGraw Hill: New York.

Jawadekar, W.S., *Management Information System*, Tata McGraw Hill Publishing Company Ltd.

Kanter, J., *Management Oriented MIS*, Prentice Hall Inc: Englewood-Cliffs.

Zani, W.S., “A Blue Print for MIS”, *Harvard Business Review*.

UNIT 6: INFORMATION SYSTEMS ECONOMICS AND SECURITY

Structure

- 6.1 Introduction: Growing Need of Information
- 6.2 Objectives
- 6.3 Data, Information and Knowledge
- 6.4 Value and Cost of Information
- 6.5 Ethics in Information Society and Right to Information
- 6.6 Protecting Computer Resources and Disaster Recovery
- 6.7 Information Systems: Success and Failure
- 6.8 Summary
- 6.9 Unit End Exercises
- 6.10 References and Suggested Further Readings.

6.1 INTRODUCTION: GROWING NEED OF INFORMATION

Maintaining records for use, as and when, has been in practice right from the ancient times. Initially people were required to keep record of their possession so that governing body could take tax from them. Government and traders kept records to plan their activities better. With the industrialization, it became necessary for industry owners and managers to keep records of inventory and accounts. Management needed more information for internal decisions. Investors, on the other hand, needed information about the organizations, its soundness, and health. In spite of the technological developments, there is an ever-increasing demand for more and more information.

Information and Information Technology have become a strategic necessity. The business environment has become much more competitive. It has become mandatory on the part of the organizations to make full use of information with the help of technology to service. In an organization, information and technology exist in the form of an information system. Whenever, an external or internal demand is anticipated or felt by the organization, information system (IS) helps the organization to plan critical response activities. The individuals are supported by IS to fulfill their roles. The organizations use IS to track consumer behavior. With the availability of technology, even in the form of hand held smart devices, it has become possible to disseminate information at any time in case organization has IS in place. The organization may take proactive measures rather than firefighting measures based on the information provided.

Typical scenario in an organization is closer to what is shown below in Figure 6.1

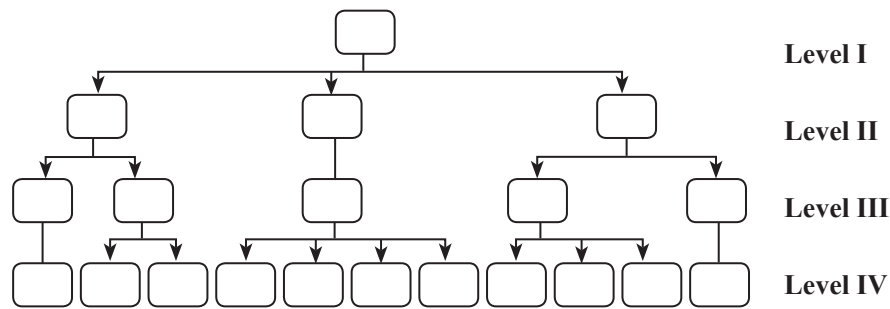


Figure 6.1: An organization as a hierarchical unit and hierarchical communication channels

At Level I is the owner i.e. Chief of the organization or top level management, Level II has middle management, Level III is lower level management and at level IV are the workers.

With increased size of organization, the volume of data generated within the organization also increased and IS became a necessity for rational decision-making by providing information at all levels by extracting from the pile of data.

6.2 OBJECTIVES

After reading this unit, you should be able to:

- Define data, information, knowledge and relationship between them;
- Explain the concepts like cost and value of information;
- Illustrate the main reasons for success and failure of an MIS;
- Describe ethical issues involved in information society and right to information; and
- Explain the need and mechanism to protect and safeguard computer resources from unauthorized access, computer viruses, and cybercrimes.

6.3 DATA – INFORMATION - KNOWLEDGE

Data is raw material with which we start and information is the finished product. For example, look at the following links:

1234	5000.00
2345	7000.00
3456	4500.00
2571	8000.00

Above lines just contain data as there is no meaning associated. Let us now put the data in the proper context as follows:

Account Number	Money Withdrawn on a day
1234	Rs. 5000.00
2345	Rs. 7000.00
3456	Rs. 4500.00
2571	Rs. 8000.00
	Total Rs.24500.00

The data is now usable and we can process it to extract information such as the amount withdrawn from account number 1234 is Rs. 5000.00. We can consolidate the data and extract the information that 24500.00 Rs. were withdrawn on the day.

Information has been defined as

Data that have been put into a meaningful and useful context and communicated to a recipient who uses it to make decisions it reduces uncertainty, reveals additional alternatives or helps eliminate irrelevant or poor ones.

Returning back to our example, the bank manager may decide the amount of required cash based on the information of total money withdrawn.

The information makes a person more knowledgeable. Knowledge is an awareness and understanding of a set of information that help decision-making. Knowledge makes a person wise. The sequence is the following – Data is processed to get Information; Information makes a person Knowledgeable, Knowledge so created adds to the Wisdom (abbreviated as DIKW). This is the relationship between DIKW.

The information should have certain characteristics to be valuable to its recipient. These characteristics vary from being accurate to secure. If information is not accurate, the decision maker may not rely on the information. The situation becomes worse if the recipient of the information is not aware of its inaccuracy. The decision maker may use inaccurate information assuming it to be accurate. The following is a comprehensive list of desired characteristics:

1. **Accurate:** The information should be accurate and error free. The information may be inaccurate due to incorrect data that has been used to generate information. The data may be inaccurate due to human error. This is commonly referred to as garbage-in-garbage-out (GIGO).
2. **Complete:** The information must be complete. The information should not have been filtered that presents a biased picture to the recipient. Let us say, salespersons of organizations are reporting sales information to the sales manager. The sales they make for the month of July are exceptionally low. They delete this information from their report whereas the sales manager might be interested in July sales just as much as in other month's sales. He might even be aware of the reasons for the low sales and might be planning to boost sales in July. The incomplete information may be useless for him.
3. **Economical:** We all understand that information has an associated cost and it is expected to be beneficial for the recipient. The benefit must be much greater than the cost.
4. **Flexibility:** Let us understand flexibility through an example. In a bank, the bank manager would like to know the total amount withdrawn and deposited through transactions, loans given and recovered by the bank. A client would like to check the total money he withdraws from his account and his present balance. The information that the bank

possesses should be flexible enough to present different views of data to different people.

5. **Reliable and verifiable:** Information is said to be reliable if one can depend on it. Both, data and information should be reliable. In case, there is any doubt or the user wants to be absolutely sure, he might like to verify.
6. **Relevant:** This Characteristic is self-explanatory. No one would like to have irrelevant data or information.
7. **Simple:** The information must be presented in proper format to make it simple for user. Too much information may result in information overload. The user may not be able to extract important information.
8. **Timely:** The information may lose its value if it is not received in a timely manner. Imagine reading yesterday's newspaper today.
9. **Accessible and secure:** The information should be easily accessible to authorized persons. At the same time, the information should be secure from unauthorized users.

To summarize, information is the result or product of processing data as depicted below.



Data life cycle:

We can think of data having their own life cycle namely, data generation, data manipulation, transmission of data (and communication of information), storing/retrieving, archiving/purging and reproduction of data.

The generation of data could take place internally and/or externally. This data has to be captured by recording of data from an event or occurrence in some form such as sales slips, personnel forms, purchase order etc.

Stored data would have to be retrieved by searching out and gaining access to specific data elements from the medium where it is stored.

Retrieved data may be converted or reproduced to different form or presentation format by way of documents reports etc.

Data are also constantly transported to the user in processed form. It is transferred to storage from the source, then processed and passed on to the user, who again returns it to storage after working on it, which becomes available for further retrieval.

The randomly accumulated data has to be sorted and classified to reveal appropriate information. For example, sales data can be classified product-wise, territory-wise, salesperson-wise etc. Such a classification will give the sales data more meaning.

Processing of data might entail quite a bit of manipulation and calculations involving addition, subtraction, multiplication, division etc. based on certain formulae. Computations might have to be performed for deriving employee's pay, customer's bill, financial ratios, hospital bills, insurance claim etc. Management science/operational research models might be used

for determining optional product mix, aggregate planning, and economic order quantity determination.

Data stored must be utilized on some occasion by some one at some point of time; otherwise there is no point in storing it. When data is finally put in a usable form it can be retrieved and turned into information at appropriate time for decision-making.

It is important to destroy data following its evaluation for use in future. Destruction/purging of data records may be accomplished on a purely routine basis following one time use or may occur in review of old records. This is done by following the organization's policy. Destruction is the terminal stage or the end of the data life cycle. The data life cycle is portrayed in Figure 6.2:

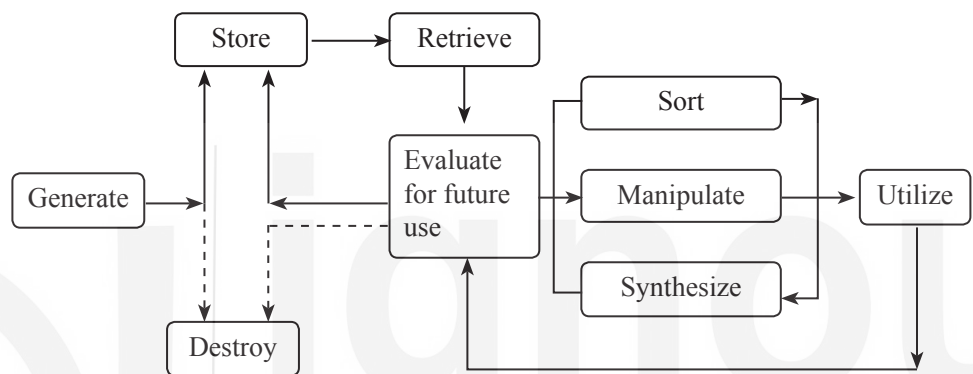


Figure 6.2: Data Life Cycle

Broken lines in above Figure indicate that unusable data should be destroyed/purged.

6.4 VALUE AND COST OF INFORMATION

The value of information is measured in terms of benefits to the organization. The benefits may be tangible that can be easily quantified. For example, 5% increase in sales is a tangible benefit, which corresponds to, say, Rs. 50,000. If the cost of the information that led to this additional profit is Rs. 20,000 then the value of the information is Rs. 30,000. Sometimes, the benefits may be intangible and cannot be easily quantified. For instance, the information may help consumers to connect to a company better. The employees may feel respected in an organization if more information is shared with them. In both the cases, the attrition rate will decrease and the corresponding benefit cannot be directly measured in terms of financial benefit to the organization.

For using information to its advantage, organization develops an information system. However before developing the system, a cost/benefit analysis is done to figure out net benefit of the system. There are many methods to assess value of information system. The most commons are explained below.

1. Cost-benefit analysis:

IT projects require investments in advance, before getting its outcomes. Many times it becomes difficult for organizations to take decisions on such investments. Over the years, however, it is clear now about the advantages of information and hence the need for the IS. Now a days

IT projects are less risky provided all resources including the human resource are well planned and put in place.

The Cost-benefit analysis method is comprehensive and explicit in understanding the investment and gain out of it hence it is covered in detail.

Identification of IT Costs

Earlier IT cost used to be so hard to estimate because of excessive overhead. With the advancements in IT and availability of human resources in much easier way than it used to be earlier the task has become uncomplicated. There are options available for investments on IT. Third party resources available, also through internet in terms of Cloud services, has helped the organizations in setting up IT infrastructure in simple way. One may go for establishments in phases based on it's preparedness of taking up the load and usability. The Table 6.1 below gives the cost involved for the organization which decides to establish its own IT infrastructure i.e. without hiring.

(a) Direct Cost

Table 6.1: Cost involved in IT

Environmental operating cost	Various items detail
Hardware cost	Servers Computers and Printers Backup devices Network and security equipments
Software cost	Operating system RDBMS Networking and security software
Installation and configuration costs	Network wiring and configuration Security software configuration In-house customizing time Re-engineering business process
Overheads	Electricity, UPS, air-conditioning, printing paper, printer toner/ cartridges, disks and backup devices
Training cost	Cost for preparing the workforce by training and re-training
Maintenance cost	Yearly service contract for hardware Software upgrades, and maintenance fee

(b) Indirect Human Costs

Indirect human cost is more significant than direct cost and it is very illusive in nature. Following is the taxonomy of indirect human costs:

- Management Time

- Management effort and dedication
 - Employee Training, re-training and human resource development
 - Management Resources
 - Personnel Issues
 - Cost of ownership
 - Employee Time
 - Employee Motivation
- (c) **Indirect Organizational Costs**
 - Losses in productivity
 - Organizational Productivity
 - Strains on Organizational Resources
 - Opportunity Cost and Risk
 - Business Process Reengineering and integration of services
 - Covert Resistance

Identification of Benefits

The following are the potential benefits of an IT system:

- Reduced Head Count
- Reduced manufacturing cost
- Reduced inventory cost
- Reduced down time
- Better quality control
- Additional new customers
- Increased sales from existing customers
- Better image of the Organization
- Higher employee morale
- Reduced attrition rate
- The ability to recruit better employees
- Eliminate data duplicacy

This approach of doing cost-benefit analysis is known as Total Cost of Ownership (TCO). The model attempts to include all costs including direct and indirect costs of owning the information system. One can include non-business uses of a computer system as a cost factor. This model is versatile and gives you complete freedom to include any relevant cost or benefits. It has been used to assess the net benefits of owning a computer system. TCO, in many cases, has revealed weaknesses of an information system in terms of under-utilization or mismanagement of IS.

Activity A

Pick an information system that you have used and do a cost-benefit analysis. Make educated guess for cost and benefit figures.

.....
.....
.....

2. Return on Investment (ROI)

Another measure of IS value is return on investment (ROI). This method tries to quantify additional profits that are generated as a percentage of the investment in information system technology. For an example, a manufacturing firm invested 5 Crore rupees in IS and an additional benefit due to increased sales is 50 Lakh rupees, then the return on investment is

$$\frac{50,000}{5,00,000} = \frac{1}{10}$$

In percentage it will be $\frac{1}{10} \times 100 = 10\%$. In other words, the return on investment is 10%. The company might perform this analysis before installing the information system with an objective of determining the utility of the system. The same analysis may be performed after the installation to check the delivered benefits of the IS against expected benefits.

3. Earning Growth

Another measure of IS value is the increase in earnings growth. Let us say, an organization experienced 7% sales growth in year 2015. After installing IS, the sales growth became 15%. Assuming that no other factors affected sales, 8% increase in the sales are due to IS.

4. Market Share

Similar to earning growth one can also evaluate value of IS in terms of increased market share. Organisational growth depends on information and IS plays the role of backbone for the organization in taking decisions.

5. Customer Awareness and Satisfaction

Customer satisfaction is one of the most valued intangible benefits of an information system. In present days, when we have online services by almost all organizations, IS helps customer track status of their orders. Customer may check the stock status before he places an order. All the leading companies that conduct survey on behalf of their client company's to determine the satisfaction level of their customers.

6.5 ETHICS IN INFORMATION SOCIETY AND RIGHT TO INFORMATION

While using IS everyone is supposed to follow ethics to refrain from committing crimes by unauthorized use as unethical use could be devastating

for the organization. To control such unauthorized use, countries have predefined rules to be followed mandatorily. India being a leader in IT industry, has IT Act which came up in the year 2000. It is widely known as IT Act 2000. Further amendments have been made for better control. Use of social media using the digital technology has also seen tremendous increase in near past. Following information shows the number of social media users only in India, in early 2021 –

WhatsApp users	- 53 Crores (530 Millions)
YouTube users	- 44.8Crores (448 Millions)
FaceBook users	- 41 Crores (410 Millions)
Instagram users	- 21 Crores (210 Millions)
Twitter users	- 1.75 Crores (17.5 Millions).

With the increase in use of IT, whenever felt necessary the Government of India issues guidelines through modifications in the IT Act. In February 2021, by modifying the IT Rules 2000 and 2011, Intermediary Guidelines and Digital Media Ethics Code were notified. As per the Act, if any person, without permission of the owner or any other person who is in charge of a computer, computer system or computer network -

- accesses or secures access to such computer, computer system or computer network.
- downloads, copies or extracts any data, computer data base or information from such computer, computer system or computer network including information or data held or stored in any removable storage medium
- introduces or causes to be introduced any computer contaminant or computer virus into any computer, computer system or computer network.
- damages or causes to be damaged any computer, computer system or computer network, data, computer data base or any other programs residing in such computer, computer system or computer network.
- disrupts or causes disruption of any computer, computer system or computer network
- denies or causes the denial of access to any person authorized to access any computer, computer system or computer network by any means
- provides any assistance to any person to facilitate access to a computer, computer system or computer network in contravention of the provisions of this Act, rules or regulations made there under
- charges the services availed of by a person to the account of another person by tampering with or manipulating any computer, computer system, or computer network then the person shall be liable to pay damages by way of compensation to the person so affected.

With the code of conduct as above, if a person accesses some data without proper authorization then the person is held responsible and could be punished. The information system and their impact must be audited and users must be aware of the risks involved.

Privacy is an important social issue involved in information society. Privacy deals with the collection and use or misuse of data. Today, there is not even a single organization not using IT. Customer data is constantly being collected and stored. This data is often distributed, sold or used without our knowledge. The health sector and retail shops, for instance, have data on its clients and customers. The fundamental question is “who owns this data and information?” We know for sure that we would not like to share our medical records with others, definitely not with insurance company or our employers.

With Right to Information Act, which came up in 2005, the Government of has empowered the citizens by giving right to be informed about the activities of the Government as informed citizen is better equipped to keep necessary vigil on governance. This Act, known as RTI Act, has further amendments for better ease and control. This Act is very useful because of digitization nationwide. Internet with the continuous developments in IT and digitization has made information flow very high. Doing online business using IT i.e. ecommerce generates lots of data. IT Act has made ecommerce more adaptive and useful. Privacy of IT consumers and electronic documents created require a comprehensive mechanism to check and control.

6.6 PROTECTING COMPUTER RESOURCES AND DISASTER RECOVERY

Crimes involving illegal system access and use of computer services, network, smart devices used over internet has increased. Hackers make use of their knowledge to gain access to others systems. An intruder may alter the data or destroy the data or may even have full control of the systems making them unusable and useless just by writing a small program. To have control over such intrusions and malpractices, organizations use various methods. For personal identification use of biometrics like finger print, iris and retina scan, face recognition are popular apart from the legacy system of user-id and password. One Time Password (OTP) is also used, based on the nature of user's business. For example, a bank may use OTP for each transaction by its customers. OTP is sent to the registered mobile of the user. Unless both the transaction password and the OTP is correctly entered, banking system will not allow to move further. Depending on the nature of computer crime that an organization anticipates, it may adopt controls. Proper access mechanism should be in place with good security devices like firewall, intrusion detection and protection system etc. in terms of both hardware and software security solutions. To summarize, suggestive control guidelines are given below –

- a) Install strong user authentication and encryption capabilities on your firewall.
- b) Upgrade your software with the help of patches, which are developed by vendors whenever a security gap is found in the software.
- c) Every employee in the organization should be given unique log-in credentials i.e. user-id and password. All such credentials should be changed regularly by the individual. System should prompt users to change the password periodically ,say, after every 90 days.

- e) Separate zones should be created and maintained for different category of users. For example, an academic institutions should have separate access zones for its employees, students, and general public. Dedicated servers for applications that communicate with outside world may be deployed.
- f) Audit trail must be used to track access log. A document once created cannot be changed without leaving an audit trail. Integrated software solutions like ERP packages, for instance, maintain audit trail. In case of a crime, the audit trail is of immense help.
- g) Install good antivirus software on all systems. Mobilize users to periodically scan the systems they are using. Email attachments should be scanned before opening and downloading files from suspicious websites must be avoided.

Information system performs key functions for an organization. If for some reason, the system becomes non-functional for some time, the consequences may be unacceptable. Organizations usually have a set of emergency procedures for critical functions. In best scenario, the end user will not be able to discover the failure of regular system. Threats to the system could be power failure, data corruption, storage failure, network failure, nature's threats in the form of a fire, flood or earthquake. In addition, labour unrest or human errors may also render system unusable.

One of first steps of disaster planning is to identify threats. After identifying the threats, appropriate disaster recovery plans for computer resources should be implemented.

Hardware backup

In case of a natural disaster or due to technology failure, the hardware may become unusable. There are companies and firms that provide disaster recovery services. A company may provide a backup system that could be used in case the primary system fails. Some companies also provide data storage services. This could be used to keep a copy of data. With cloud facility over internet backup system has become very easy to maintain.

Software Backup

Software programs are precious assets of an organization that must be protected. A human error may delete a software package or a hardware failure may make it inaccessible. A simple strategy is to make copies of software and keep them safely. In addition, one may like to keep another copy off-site in a safe environment.

Regular periodic backup should be adopted in practice. If the data is too large, incremental backups can be taken or selected data may be backed up at regular intervals.

The smart strategy is to be in pro-active mode rather than reactive mode. It may be less expensive to plan ahead to avoid possible down time than suffer losses.

Virus Protection

Computer viruses – sometimes known as malware – come in many different forms. Till 2020, these were an incredible 350,000 new pieces of malware,

each one being a single piece of intrusive software. Viruses can find their way onto just about any device – from computers, smart phones, to network of an organization. A virus reproduces itself, usually without your permission or knowledge. In general terms, they have an infection phase where they reproduce widely and an attack phase where they do whatever damage they are programmed to do (if any). Another way of looking at viruses is to consider them to be programs written to create copies of them-selves. These programs attach these copies onto host programs (infecting these programs). When one of these hosts is executed, the virus code (which was attached to the host) executes, and links copies of it-self to even more hosts. Many viruses do unpleasant things such as deleting files or changing random data on your disk, simulating typos or merely slowing your system down. Some viruses infect other programs each time they are executed; other viruses infect only upon a certain trigger. This trigger could be anything; a day or time, an external event on your system etc. Viruses often delay revealing their presence by launching their attack only after they have had ample opportunity to spread. This means the attack could be delayed for days, weeks, months, or even years after the initial infection. However, viruses are only one way your data can be damaged. You must be prepared for all threats; many of which are more likely to strike than viruses such as storage failure due to hardware problem. All software and data reside in storage. There are many other threats to your software and data that are much more likely to harm you than viruses. There are a large number of virus types. The most common computer viruses are categorized as below –

Macro Virus - This type of virus is normally found in Microsoft Word and Excel files. This virus increases the size of files when it infects them attaching its own generated codes. Once a macro virus infects a file, it can easily spread to other computers when that file is shared, for example via email or copied through other storage media.

Boot Sector Virus – This is the oldest virus affecting the startup or ‘boot’ process. Back in 90s, these viruses were spread through floppy disks. Nowadays, they attach themselves to emails or USB sticks. If your computer catches one of these, you’ll need to carry out a full system reformat.

Trojan Horses - Used by cyber-criminals, Trojan horses are disguised as normal programs, tempting you to install them on your computer. Once installed, the viruses gain access to your computer’s files and capture your private data – think passwords and online banking information. This can then be used by hackers to make online purchases with your bank account or expose your private information.

Overwrite Virus - These viruses typically take over a file and wipe the original code without you even knowing it. Once deleted, the original files cannot be recovered and the data is lost. They often spread through emails via attachments or through file downloads on the internet.

Browser Hijacker – This virus hijacks your internet searches and redirects you to pages you didn’t even want to visit. While not as harmful as other types of viruses – they trick you rather than steal from you – browser hijackers are still an annoying problem since they significantly lower your system when surfing the internet.

Web Scripting Virus -Web scripting viruses are very clever little bugs that blend into the background of popular websites – usually social media platforms. They disguise themselves as normal links, tempting you to click on them. Once you click, they steal your information. They can send spam to your system and damage your data, and can spread faster than most viruses.

Polymorphic Virus - A polymorphic virus modifies whenever it replicates. This makes it hard for most anti-virus programs to keep up. Once it's found its way onto your computer – usually through an email attachment or a download from a suspicious website – it has free reign to delete your files, steal your data, and generally sabotage your system.

Resident Virus – This is the most common type of virus. It finds way into your computer's memory, completely uninvited, and resides there. Resident viruses can come from email attachments, infected download or shared files.

Multipartite Virus - They usually spread through .exe files. – so programs like Word and Excel. These viruses eat up your virtual memory. You start getting messages like 'your computer has low virtual memory' resulting a sudden slowing down of your computer.

Spacefiller Virus – This is also known as “cavity virus”. These viruses find the empty spaces in a program code and add their own code. This way, they don't alter the size of files but corrupt the program itself. They are, however, very rarely found.

Cyber Crime and Cyber Security

With internet overpowering the huge base of population globally, cyber crime has emerged as a big threat to individual and organizations. Cyber crime is a criminal activity that either targets or uses a computer, a computer network or a networked device. Such activities are carried out by individuals or organizations. Types of cyber crime may involve fraud with individual's emails, phishing and sending spam mails, personal or corporate data stealing and using, cyber-extortion and ransom-ware attacks, cyber-spying to access government or company data etc. Cybercriminals use viruses and malware to commit such crimes including the Denial-of-Service (DoS) attack. If such a situation prevails then total business comes to halt as credentials are lost and the system goes in control of the cybercriminals.

Cyber security is the application of technologies, processes and controls to protect systems, networks, programs, devices and data from cyber attacks. It is to be noted that information security deals with protecting the integrity and privacy of data, both in storage and in transit, cyber security is required as a secret weapon against cyber criminals with the aim to reduce the risk of cyber attacks and protect against the unauthorized exploitation of systems, networks and other such resources. Since it is applicable to all, working online through internet, Governments have made it mandatory for the organizations in their countries to have sufficient control mechanisms deployed. In India, Companies Rules 2014 framed under the Companies Act 2013 requires companies to ensure that electronic records and security systems are secure from unauthorized access and tampering. Similarly Reserve Bank of India (RBI) provides guidelines for banks to secure banking businesses. All the banks are supposed to mandatorily follow the ISO/IEC

27001 and ISO/IEC 27002 standards for ensuring adequate protection of critical functions and processes. Similarly, SEBI requires stock exchanges, depositories and clearing corporations to follow standards, such as ISO/IEC 27001, ISO/IEC 27002 and COBIT 5. Cybercrime Cells are also available for individuals and organizations to lodge complaints at governmental level.

There are many organizations which provide cyber security as a service. They provide complete solutions to mitigate and resolve issues related to cybercrimes.

CERT –In

In accordance with the IT Rules, the Computer Emergency Response Team of India (CERT-In) came into existence in 2004 in the Ministry of Electronics and Information Technology (MEIT), Government of India. It is operational since January 2004. Cert-In is the nodal agency responsible for collection, analysis, and dissemination of information on cyber incidents and taking emergency measures to contain such incidents. It responds to computer security incidents as and when they occur. CERT-In has been designated to serve as the national agency to perform the following functions in the area of cyber security:

- Collection, analysis and dissemination of information on cyber incidents,
- Forecast and alerts of cyber security incidents,
- Emergency measures for handling cyber security incidents,
- Coordination of cyber incident response activities,
- Issue guidelines, advisories, vulnerability notes and whitepapers relating to information security practices, procedures, prevention, response and reporting of cyber incidents,
- Such other functions relating to cyber security as may be prescribed.

(Source <https://www.cert-in.org.in>, <https://www.lexology.com>)

6.7 INFORMATION SYSTEM: SUCCESS AND FAILURE

An Information System is developed to assist management in problem specific decision-making. IS development and deployment project may succeed or it may fail. A project is a success if it is completed within time and budget and delivers all as decided in beginning. It must meet needs of its users and organization. Over a period of time, the following main success and failure factors have been identified. These days integrated software modules are available which are required to be customized to have the software in place. The software, so customized, may be either installed in in-house serves or on cloud servers. In cloud servers and storage, organization does not need to have servers and storage in its premises. Access to the IS is through internet hence only computer/laptop/smart devices are required to use the IS. There are many factors that contribute to the success of a project. They are detailed below.

1. The project scope should be stable and well understood. If the scope of the project changes during the development or customization of the software, the project is likely to suffer in terms of quality, schedule and budget overrun.
2. An MIS project that aims at re-engineering the business processes of an organization faces major challenge. Such projects are high-risk but at the same time have high potential for major benefits.
3. The technology development platform and development language exposure are other critical factors. Sometimes, the technology may be new and the team may have difficulty using the technology. The platform and language newness may also create trouble for the team.
4. Support from the management is vital for the success of the project. If management loses interest in the project, budget may be cut, key people may be moved to another project or the moral support required by the team may become non-existent.
5. The objective of MIS must be in tune with the objective of the organization. For instance, objective of an organization is to cut cost. An MIS that aims to handle financial transactions of the company is not in tune with the company's objective.
6. The system should be user friendly and the response time should be reasonable so that the user feels good to work on the system. User should never be over-powered by the system.
7. MIS should be developed with a clear objective that must be documented before the development commences. The objective must be identified with the help of all stakeholders. The system analyst must interview concerned people to establish their needs. Inputs should be consolidated and presentation should be made in front of stakeholders and the development team. This should go till the user requirements are fully covered and understood i.e. till the objective becomes clear.
8. An important aspect of an MIS is data. Data policy such as what data will be included, who will provide the data, who will validate the data, who will integrate the data and how will the data retire should be clearly stated. These seemingly trivial issues can turn a success into a failure.
9. A quality control plan must be in place to ensure quality of specification and quality of conformance. If required, a third party may be engaged for this purpose.
10. The concerned people should be adequately trained on the new system. This skill development process should go on periodically to overcome gaps due to shifting of human resources or because of changes made in IS.
11. The system should be properly and explicitly documented so that attrition does not affect the project or MIS adversely.

6.8 SUMMARY

In this unit, you have been introduced to basic concept of data, information, and knowledge. You have also learnt the growing need for information in

the society, ethical issues and right to information. The life cycle of data was introduced. The information has a cost and value associated with it which was discussed in some detail. Information System should be protected and secured hence concepts of unauthorized access, cybercrime and cyber security was discussed. Steps for disaster recovery and threats of virus to mitigate and resolve security issues were discussed. An information system depends on many factors for its success. We also discussed the main reasons for success and failure of an MIS.

6.9 UNIT END EXERCISES

1. Define data and information. What is the difference between the two?
2. What are the main characteristics of information? Why do we require in organizations?
4. What are various ways of assessing the value of information? Explain each method briefly?
5. Write a detailed note on data life cycle.
6. What are the main reasons for success and failure of an information system?
7. Does an organization have a right to collect and share information without the permission of person concerned? What are the ethical issues involved in information society.
8. Why should every organization have a disaster recovery plan to protect itself?
9. Write a brief note on virus threat and a protection strategy.
10. Write a brief note of cyber security and cyber crimes.

6.10 REFERENCES AND SUGGESTED FURTHER READINGS

Bishop Matt(2002), *Computer Security: Art and Science*, Addison-Wesley Pub Co; 1st edition.

Davis, G.B., 2000, *MIS Conceptual Foundations, Structure and Development*, McGraw Hill: New York.

Jawadekar, W.S., 5th Edition, *Management Information System*, Tata McGraw Hill Education – Europe

Kanter, J., *Management Oriented MIS*, Prentice Hall Inc: Englewood-Cliffs.

Pfleeger Charles P & Pfleeger Shari L., (2002), *Security in Computing, third edition*, Prentice Hall PTR.

Zani, W.S., “A Blue Print for MIS”, *Harvard Business Review*.

<https://www.kaspersky.co.in> site accessed on June 9, 2021

<https://www.cert-in.org.in> site accessed on June 9, 2021

<https://www.meity.gov.in/content/icert> site accessed on June 10, 2021

<https://www.lexology.com> site accessed on June 10, 2021

<http://www.cknow.com/vtutor/vtintro.htm> site accessed on June 10, 2021

UNIT 7: TRANSACTION PROCESSING SYSTEMS, DSS, AND EIS

Structure

- 1.1 Introduction
- 1.2 Objectives
- 1.3 Information from Data
- 1.4 Information in Managerial Functions
- 1.5 Information Systems in Management
- 1.6 Programmed and Non-Programmed Decisions
- 7.7 Environmental and Competitive Information System
- 7.8 Information Systems in Functional Areas
 - 7.8.1 Marketing
 - 7.8.2 Finance
 - 7.8.3 Human Resource Management
- 7.9 Transaction Processing System
 - 7.9.1 Human Resource Management Sub-System
 - 7.9.2 Marketing Management Sub-System
- 7.10 DSS, and EIS
 - 7.10.1 Decision Making Process
 - 7.10.2 Decision Support System
 - 7.10.3 Executive Information System
 - 7.10.4 Difference between MIS, DSS and EIS
- 7.11 Summary
- 7.12 Unit End Exercises
- 7.13 References and Suggested Further Reading

7.1 INTRODUCTION

We know the importance of knowledge. It is considered to be a supreme source for all activities. Well-appreciated strategic decisions are the outcome of good knowledge. Acquirement of knowledge at one point of time and continuous updating keeps a person fit for business survival in the competitive environment. Learning from personal experience is one way of gaining knowledge. Knowledge gained through sources like education, interaction with others and other means is comparatively faster. Information enriches the knowledge. Role of information and knowledge in managerial decision-making is inevitable in nature. In managerial functions, knowledge integrates the activities of different departments and enables the decision maker to take right decisions. Further information enriches the managers about the latest scenario on the vital areas of functional management like Finance, Marketing, Human Resource and Operations. Information is not only the foundation for any business; it also plays the role of bricks

in building up the business decisions. The generation and dissimulation of information should be a continuous process in the organization.

In this unit you will get an overview of basic information systems with focus on functional areas. The coming two units will deal with HR & Marketing Management and Operations & Financial Management respectively.

7.2 OBJECTIVES

After reading this unit, you should be able to

- Describe the importance of information in managerial functions;
- Identify the role of information;
- Identify and appreciate the basic information systems;
- Diagnose the situation for development of information system; and
- Attribute the role of information systems in functional areas of management.

7.3 INFORMATION FROM DATA

In simple terms, information is defined as the processed form of statistical data. It is important to note that information and data are different to connote. But they are interrelated. The raw data collected about any phenomenon may be used to generate the required information.

While processing the data for the information, one needs to keep the following in mind-

- Data should be accurate
- Data should be relevant
- Data should be complete
- Data should be time sensitive.

Information processing from the raw data can be represented as information funnel. Procurement of correct information in right time is one of the important aspects for business decisions. Having the right information in hand, business organization can be proactive in its activities to maximize the business profits. The following schematic chart indicates the position of the information among other relevant characteristics like data, knowledge, and wisdom.

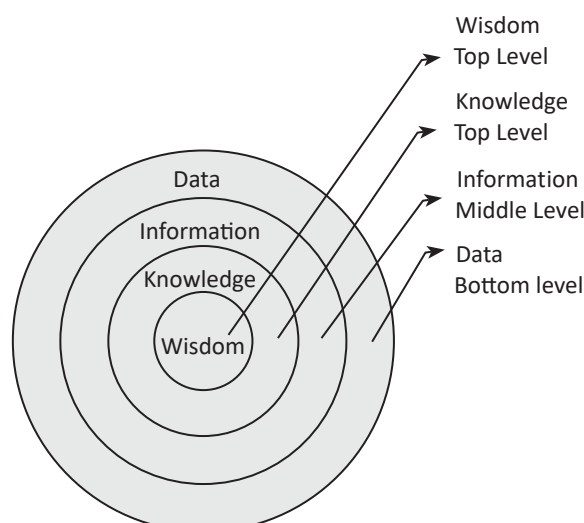


Figure 7.1: Position of the Information

Information from data is generated by going in to different stages in the process. They could be stated as:

- Summarizing
- Data trimming /mining
- Presentation and
- Indexing

Data collected for want of information are usually available in the form of questionnaire, schedules, note, or the official unclassified formats. There could be secondary data as well from the web. The first step towards the analysis is summarizing i.e. bringing the data into the required form of tables or arrays so that further treatment of data is possible. Grouping of the data under different category will enable the analyst to identify a specific tool for analysis. The computerized data processing for summarizing the data through selected software expedites the process.

Data trimming is the condensation or consolidation of statistical data for retrieving information. There are specific statistical data reduction tools for trimming the data. Depending upon the nature of the data and the objectives of the study, the tools are decided. For example, with the application of statistical packages related to the inventory problems, it is possible to derive information, say, for re-ordering of items with sufficient lead time so that production does not get stopped because un-availability of an items.

The results obtained after data trimming need to be related to the objectives for the meaningful representation. The process of depicting the statistical results in the desired form (charts, diagrams, schematic representation etc.,) is called presentation. The presented results will reveal the information in a better way. Information, as an object could be derived from the presented results.

Indexing or codification of the results is required for further processing of information like storage and retrieval, modification, converting into electronic mode for communication etc., The codification is the base for presenting the information on web pages. In the Internet search the codes are recognized by the computer at a greater speed to retrieve them from the data dictionary or catalog. The big computer databases are designed and used for data retrieval through proper indexing of the data. Thus raw data collected from the field becomes information in the final form.

Types of Information

Information retrieved from raw data could be classified in different categories as:

1. Facts
2. Assumptions
3. Collateral information
4. Fragmentary information
5. Irrelevant information

Facts are the accurate information. The information generated with the assumption of mathematical models is called reliable information without

any uncertainty. In some statistical analysis certain amount of probability is used to predict the required results. Such information may not be 100% true. This type of information is called information based on assumptions. The other type of information is called collateral and fragmentary information. This type of information cannot be directly retrieved but is a parallel source of information. It is not considered to be reliable. Such sources need to be investigated for acceptance.

The manager responsible for taking decisions requires sufficient information from the relevant records available with the organization and if need be from web also. The term information means collection of facts from which business decisions can be drawn. Facts and figures collected in this connection will enable the company to go for a valid decision.

7.4 INFORMATION IN MANAGERIAL FUNCTIONS

With the representation of Antony's structure, the data available at different functional areas of management could be presented as follows:

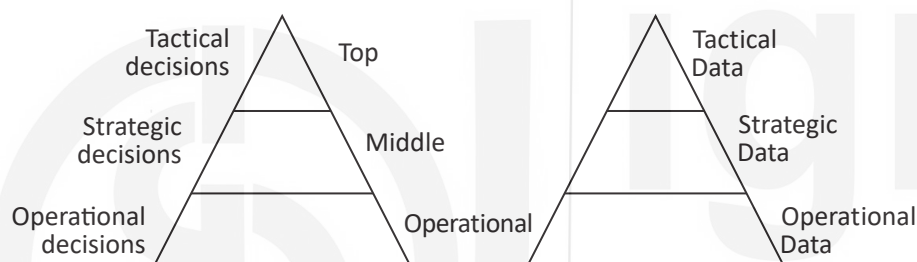


Figure 7.2: Data Available at Different Functional Areas

At operational level, organization gets in contact with public for information procurement and dissemination. All types of inputs available from various sources are gathered at this level. The routine office work like maintaining inward register, maintaining of public relations, understanding the external requirement etc are mostly done at this level. Though no decision making process is carried out, proper accumulation of data has to be done at this level for further processing. Hence it is sufficient to have a trained manpower that can act as per the computer-assisted instructions. At middle level of management, the input obtained from internal and external sources of information will be processed for strategic decisions. This level consists of middle level managers who can contribute significantly for the development of organization. Different options available for each of the objectives are evaluated carefully using the scientific tools of analysis and outcomes are reported. The role of middle level decision process is considered to be very crucial because the output received from this level is taken as input for tactical decision-making. The highest level in this hierarchy is the tactical decisions. This is the decision making group which should ultimately be supported by the online information. To highlight the information needs of the organization; the sources are classified as internal and external. In general, the internal data sources for any organization, irrespective of its nature, are Accounting control, Plans and budgets, Payroll, HR, Inventory, Sales forecasting, Purchase, Transportation and distribution, Queuing,

Operations, R & D etc. Similarly the external sources could be Economic factors, Environmental factors, Market trend/constraints, Labor oriented problems etc.

Educating the managers of the companies to understand the information generation, treatment of information and applications in real decision-making is very essential. The task of information generation, process of information, management of information and timely communication or dissimilation is interrelated and hence a systematic approach is required. As the business decisions are based on the available information, adequate care should be taken to have the right information.

The information perspective in an organization is viewed categorically for managers as well as for the organization. The managers use the information as an object for managerial decisions, as a construct to build up the organization and tools to gain explicit and tacit knowledge. But for the point of view of the management, information should be treated as a resource, as perception pattern for the development, as a commodity to be appreciated and as constituent force to meet the internal and external changes. Figure 9.3 shows the links among the perspectives on information in an organization and as a manager.

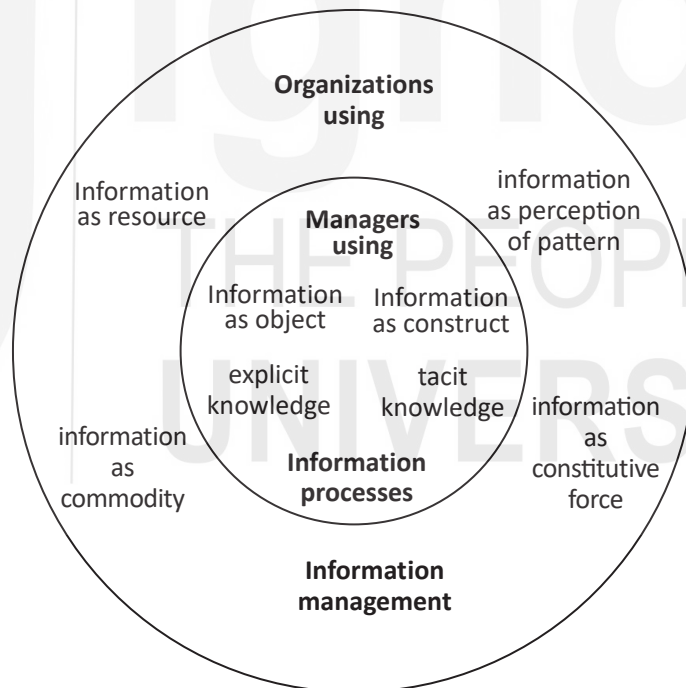
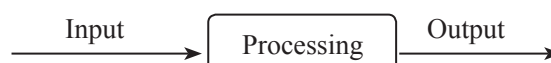


Figure 7.3: The information environment in organizations

7.5 INFORMATION SYSTEMS IN MANAGEMENT

Information handling in an organization should be a systematic process. Generally systems are created to solve problems. One can think of the systems approach as an organized way of dealing with a problem. A collection of components that work together to realize some objective forms a system. Basically there are three major components in every system, namely input, processing, and output.



In a system the different components are connected with each other and they are interdependent. For example, Human body represents a complete natural system. We all are bound by many natural systems such as political system, economic system, educational system, health and hospital system and so forth. The objectives of the system demand that some output is produced as a result of processing the suitable inputs.

In business information processing system, input is received as data and instructions, processing is done under the defined instructions and then output is generated. The inputs to the system may be from a data file, Internet source, or manual feed. Further the system interacts with human or computers to support the operational and managerial decision making of an organization. The managers should be well aware of information systems because:

- The amount to be spent on the information system should have positive effect on profit making,
- It increases business awareness, promotes business re-engineering and enables total quality management to the managers,
- If the managers are not acquainted with the information systems, significant amount will be wasted on automating ineffective process or at the extreme, companies can fail as a result of poor management.

The organization expands in the direction of information expansion. Flexibilities in information handling help the managers to take advantage of all business opportunities irrespective of small or large-scale enterprises. Some of the important implications about the IS deployment and management on information are as follows:

- 1) Information system will contribute substantially to the managers in their effective decision making to achieve the organizational objectives,
- 2) Information system will enrich the organization to meet the competitive challenges,
- 3) Adequate knowledge from the past experience will help the Managers to integrate the business strategy,
- 4) Information system will provide sufficient awareness about the external components to react accordingly,
- 5) Information system is required to take right decisions in right time.

The effectiveness of information management can be measured by the extent of knowledge creation or innovation in organizations. The process approach to information management supports the integration of business strategy and information. Some of the important implications about information management are as follows:

- 1) If information is conceptualized as a process it can be integrated into strategy formation
- 2) Information management should account the information as an object and as a construct

- 3) It has a role in enhancing the information capabilities of individuals in organizations
- 4) Almost all managerial activities have an information component. So information management practice should be responsive to the needs of managers as information users and as information producers or knowledge creators

The importance of information in management is further emphasized by the fact that much of management is primarily decision making. While there are several views of what constitute management, the generally accepted planning, organizing, coordinating, directing, and control are all concerned with decision-making. Several functional areas of management viz. Personnel, Marketing, Production, Finance and Services will be significantly influenced by the information systems that are to be implemented in an organization. Information once acquired need to be communicated promptly to various departments as per their requirement. The free exchange of information in an organization is called information flow. Information flow integrates various levels of management discussed earlier. Operational decisions deals with a large volume of internal data. The middle management is concerned with medium range or tactical decisions required for much less information. The top management being concerned with long term or strategic decisions requires for just a few vital internal information but a lot of external information. The successful information system should take into account such a pattern of information needs by the management. Typical example of the three levels of information among the functional areas of management can be represented with the help of a pyramid structure as given in Figure 7.4.

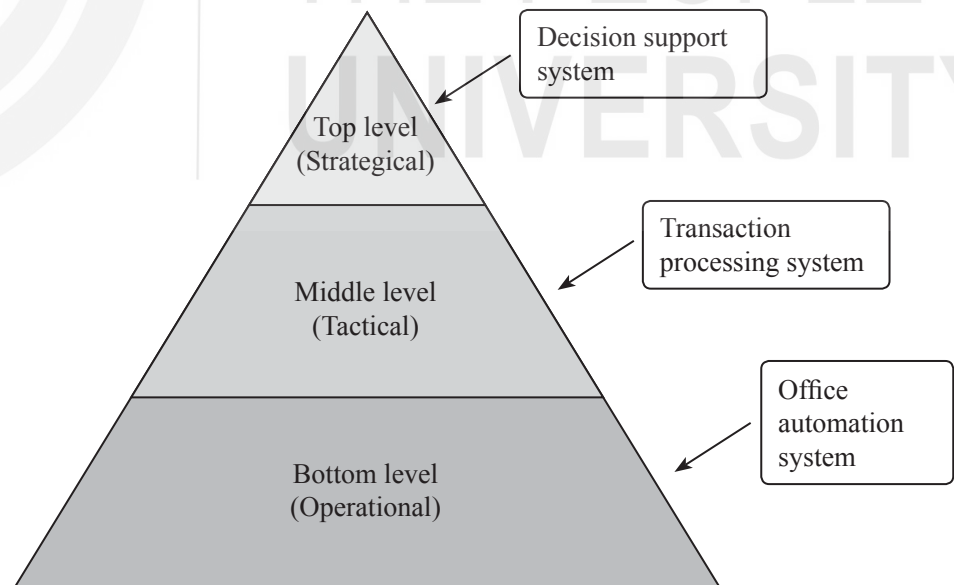


Figure 7.4: The information systems and Management levels

All the four major areas of management are inter-related. Creating a master database comprising of the activities of the organization irrespective of the specific area is a must. It should be possible to retrieve the data for any further information. This kind of approach is called integrated approach. The master database created should be flexible to have the following features:

- 1) Distribute and share the data on network
- 2) Normalized to accept the unique characteristics,
- 3) Capable of automatic updating
- 4) Quick query

The following diagram depicts the creation of centralized database for application to various functional areas of management.

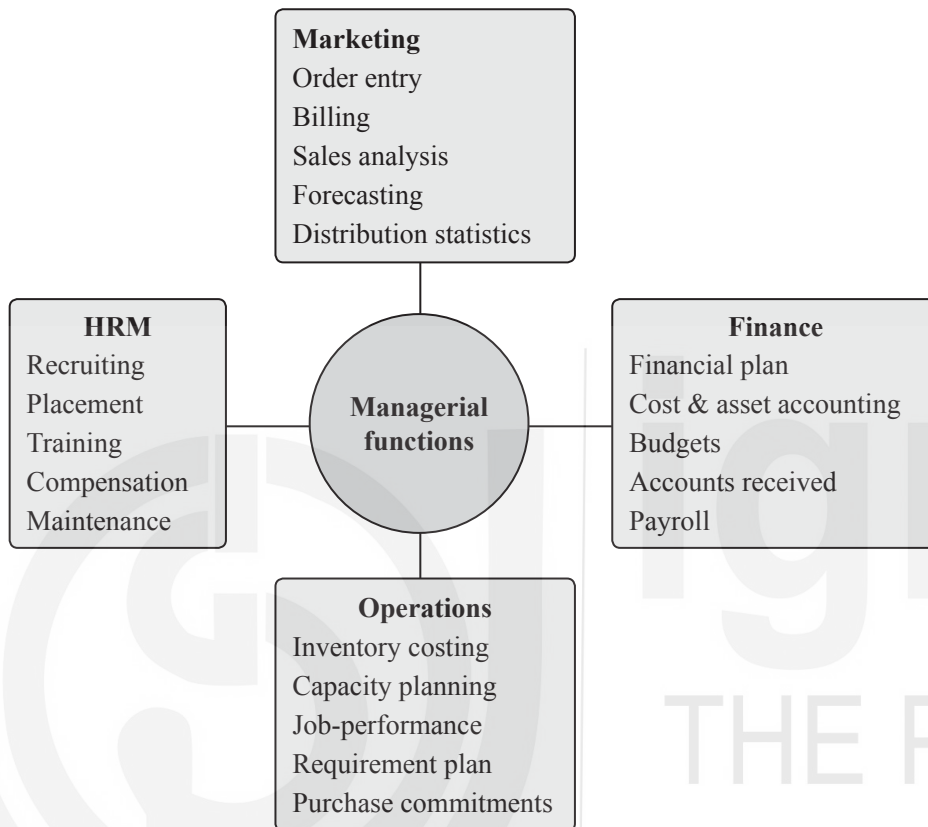


Figure 7.5: Creation of Centralized Database for Various Functional Areas Application

A typical example of role of information systems for the important areas of functional management at different level of decision-making is shown below:

Table 7.1: A typical example of role of information systems

	Production	Finance	Personnel	Marketing
Strategic	New plant Location	Alternative Financing	Welfare Policy	Competitor Survey
Tactical	Production Bottleneck	Variance Analysis	Performance Appraisal	Advertising
Operational Daily	Scheduling	Payroll	Leave Records	Sales Analysis

7.6 PROGRAMMED AND NON-PROGRAMMED DECISIONS

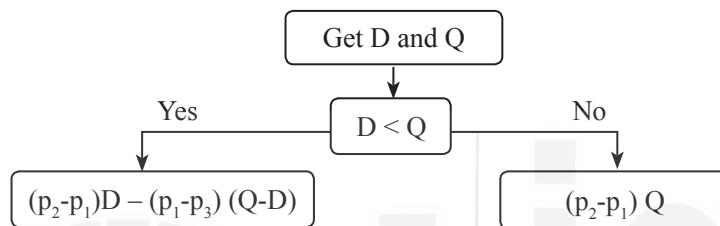
Managerial decisions are classified into programmed and non-programmed categories. There are some situations in which the analytical methods of

decision-making are possible and in some other situations one need to apply the intuitive methods of decision making. Programmed decisions are those that can easily be automated, like the determination of optimal product mix, minimum cost production schedule, optimal sequencing of machines to minimize mean flow time etc. Generally large data and a few decision rules or algorithms that use the data in an automated fashion to arrive at an optimal plan characterize such decisions. Techniques of Operations Research like Linear Programming represent a typical example of this category of decisions that use formal data and algorithms. Naturally such decisions are easily programmed. In other words they can be represented as an algorithmic procedure into unambiguous instructions whose step-by-step execution will lead the optimal result. Since these algorithms are likely to be codified in the form of a computer program and run on a digital computer, they are programmable or programmed decisions. The key to such programmability is the underlying structure of these decision situations that permit an algorithmic translation. Many of the programmed decisions may need the most challenging algorithms involving the best brains available at that moment for their solution. Nevertheless they are translatable into algorithmic procedures. Simon's predictions of the fifties have practically come true in the nineties. This is borne out by the large number of decision support systems that use data and models to automate such decisions like reorder levels in materials and production control, control limits for process re-setting etc. Information support for such programmed decisions can be designed rather easily.

Non-programmed decisions on the other hand do not lend themselves to easy automation. The model support for such programmed decisions would need more of heuristics rather than optimal algorithms. Optimization Algorithms as mentioned earlier, are formalized procedures readily implemented as a finitely terminating, computer programs with guaranteed outcome of optimal solutions, whenever they exist. Heuristics, on the contrary, generally yield near optimal if not the optimal solution but cannot guarantee optimality. The heuristics themselves may be implemented in the form of an algorithmic procedure and solved on a computer; however the decision situation may not have admitted any optimal algorithm. The interesting developments in complexity theory of the theoretical computer science also led to an interesting observation that for many interesting optimization problems, heuristics are necessary as no optimizations algorithms with reasonable estimates of computational performance are known to exist. This idea led to the heuristic problem solving approach where one gets good solutions and even optimal solutions but optimality cannot be guaranteed. Such heuristic procedures for the solution of many real-world problems in the area of scheduling and resource allocation have been found to be extremely successful.

Over the years the programmed and non-programmed decisions have also been known as structured and unstructured decisions. While designing, information systems one should keep this vital difference between the two types of decisions in mind. Flexible access to data, user friendliness, graphical user interfaces, natural language support, what-if analysis, capability, etc. characterize information support for non- programmed decisions.

Example for programmed decisions: Consider the problem faced frequently by a Small and Medium sized Enterprises (SMEs) in deciding about the size to be ordered to maximize the profit on consumable items, say, number of units of food packets. The small-scale enterprise is doing the business of trading between the wholesaler and retailer. The purchase price per unit is p_1 and the selling price is p_2 where $p_1 < p_2$. The unsold item for the season/for the day will make a loss to the dealer and hence the dealer will be able to return it at the lower price p_3 ($p_1 > p_3$). The objective is to decide about the number of units to be ordered to maximize the profit. The intuitive approach to this problem will lead to problem specific and cannot be conceptualized. It is better to think of using structured analysis or analytical approach. Assume that the number of units demanded is D and the quantity ordered is Q . The logical flow of this approach is



This is called conceptualization of the problem. Once the problem is conceptualized, we can give the mathematical function to the problem.

The mathematical formulation to find the profit to this problem with the above notations is:

$$\text{Profit} = \begin{cases} (p_2 - p_1) Q & \text{if } D \geq Q \\ (p_2 - p_1) D - (p_1 - p_3) (Q - D) & \text{if } D < Q \end{cases}$$

The model defined above could be used for any available quantities of D and Q . The advantages of this model are:

1. The system is open to anyone in the organization for application
2. Instantaneous and accurate calculation is possible
3. Enables computerization of the system
4. Flexible for review/modification.

This kind of approach is called systematic approach or programmable decisions, which is the base for the system development for retrieval of information.

Example for non-programmed decisions:

Managers are quite efficient to make better decisions as they have got a great deal of experience. Sometimes decisions need to be taken in the light of the situations and not in general. The persons who are in the field for quite some time will understand easily and the decision can be taken accordingly. For example, to commission a power plant, conducting an in house survey about the past performance of the plant is sufficient to gain adequate information instead of going for a generalized study.

In non-programmed decisions, the conclusions could be arrived through the following simple steps (Dewey):

- 1) Identify the problem
- 2) Identify the alternatives
- 3) Select the best alternative.

Though the programmed decisions are considered to be more accurate, there are some situations in which the assumptions made under this category cannot be fully practical and hence the approach of non-programmed decisions are used. If the decision maker is well experienced, no doubt the accuracy could be maintained and admit that the experience improves the problem solving.

7.7 ENVIRONMENTAL AND COMPETITIVE INFORMATION SYSTEM

The information system developed for business purposes should take into account all the conditions, which affect the business transactions. Usually the two major factors responsible for the business changes are the environmental conditions (local and global) and the comparative factors. Successful development of information systems should have been based on the deep understanding of the organizational structure and dynamics of the enterprise also. For production department the assignment of jobs to machines or machines to operators in a production environment is very important. For a service organization like hospital it may be a duty roster for nurses, doctors or other paramedical staff. Such schedules must be detailed and unambiguous reports could be produced in large number at relatively low cost. Most of the information contained in such reports is internal - to the special shop or hospital. Some organization are goal oriented, the analyst must be clear as to what information exactly need to be collected, stored and analyzed. Since every information must have a context, only operational information that ultimately has some decision-making contribution must be collected. Secondly, the information collected and processed must be consistent with the level of the organization to which it is to be presented. Organizations have not only found out the immense utility of information systems to better manage their organizations but are also feeling the potential of the competitive advantage provided by, information as a resource. The classic example of the use of information for competitive advantage is Airline Reservation System and the Frequent Flier Option introduced by Airlines the world over. These are examples that illustrate the use of information not just to run business effectively but also to open up new business vistas. Similarly festive seasons offer by Retailers and credits given to customers by them beyond certain limit of purchases could be another example.

The payroll represents a typical operational system for the financial management. Since such systems must execute accurately data pertaining to a large number of staff in a timely manner, month after month, cost based efficiency and speed would be a major consideration in the design of such a system.

Detailed Sales Analysis is a must for any sales and marketing function. This might involve a very detailed data collection and processing pertaining to

every salesman, every product over a long time span of an entire year or a quarter with even details of the region, market segment etc. Accuracy and timeliness of this effort would decide the quality of tactical and strategic decisions that may be based on summary information generated out of this data. Accuracy and desegregation would be the watchword for such detailed data collection effort.

Naturally information systems must have summary information, with provision for detailed information to provide such information. Comparative analysis shift wise, plant-wise, machine-wise, operator-wise etc. may be necessary. The watch word would be insight & analysis and not mere reporting of data Variance Analysis for the finance function would call for systems that point out deficiencies, cost over runs, budget excess by carefully matching goal or target information with summary information generated by operational data. Once again the accuracy is important; but what is more important is the clear recognition and highlighting of patterns that can help the decision maker to initiate action and bring systems under control. Timeliness is a must, as the systems must be reset before it is too late.

A tactical system under personnel function is the Performance Appraisal, which would take into account among several other things, the individual employee leave records. The details may not be that important but patterns are. Whenever they are clearly visible. Certain policy conformance may also be checked through such analysis. Since decisions based on such systems are likely to have medium to long-term impact, accuracy should not be underestimated, while keeping the priority of analysis.

Advertising Information System is another examples of tactical information system for the marketing function of management. Needless to say planning for advertisement would use much of the information generated by detailed sales analysis (an operational MIS system). But to be able to decide on the levels of advertisement, mix of advertisement, budget for advertisement etc., calls for certain policy parameters as well as environmental (outside organization) information. This point must be clearly borne in mind in designing MIS for tactical decision situations.

A strategic decision by the production function of the management is a New Plant Location. This would use much of the internal information generated by the tactical system designed to analyze the production bottleneck that is internal to the organization. But a long term decision like location of a new plant is likely to be influenced far more by environmental information like changing market, changing technology, changing fiscal and governmental policy like deregulation, tax incentive for backward area etc. Obviously strategic information systems should have a mechanism of scanning and assimilating environmental information that are likely to influence strategic decisions in a systematic way. For the finance discipline, alternative financing is a strategic decision. It will use a summary status information about the internal finances of the company including payroll, budget, overruns, variance analysis etc. but will be primarily governed by long term policy, business environment etc. that are strategically important to the survival and growth of the organization. Information support for

such strategic decision would call for substantial external information supplemented with internal financial health indicators. Decision concerning welfare policy of an organization is a strategic decision that must be faced by personnel (also known as human resource development) management. It will be tempered by internal information about staff size, their quality, and the compensation package etc. but the strategic decision will be governed by the future vision for the organization as seen by the top management as well as the labor market conditions of the environment. Being a long-term decision with major impact on the corporate health of the organization this strategic decision need to be supported by a balance blend of external and internal information. While we take primarily the decision-making role of the management in the design of the information systems, care must be taken to take into account the stark realities of managerial work-style. The diversity of managerial work, the interpersonal dynamics; politics of people, resistance to change, etc. must also be taken into account. These are issues beyond the scope of this text but must be taken into account in the actual implementation. A strategic decision for the marketing function is a survey of competition and the resulting strategy of gaining market share. While a detailed analysis of in-company sales and advertising is a necessary prerequisite, much will depend on the present and future strategies of the competitors who are external to the environment. Once again this brings to the focus the importance of external factors for strategic decision, policy likes deregulation, tax incentive for backward area etc. Obviously strategic information systems should have a mechanism of scanning and assimilating environmental information that are likely to influence strategic decisions in a systematic way.

7.8 INFORMATION SYSTEMS IN FUNCTIONAL AREAS AND DECISION-MAKING

Let us discuss some of the information systems in functional areas and their contribution to decision making.

7.8.1 Marketing Management System

The term marketing as per the American Marketing Association (AMA) is defined as the process of

- Planning and executing the conception,
- Pricing,
- Promotion and
- Distribution of ideas, goods, and services to create exchanges that satisfy individual and organizational objectives.

Marketing function facilitates the activities mostly related to -

- Forecasting and sales planning
- Market research
- Advertising
- Operation and control of information required for market.

The management of information, which provides output to the above-specified activities, is one of the important aspects in the business. To fulfill the objectives of the marketing, periodic tasks pertaining to collection of data, analysis and presentation in the form of required reports need to be performed. The integrated approach comprising the data input, generation of marketing information and supply of output in the required format is called marketing information system. It consists of people, equipment, and procedure to gather, sort, analyze, evaluate, and distribute timely, and accurate information to decision making. Marketing information system are expected to help the management in decision processing for pricing the product, packaging, new product development, product mix analysis, advertising, product promotion policy, sales strategy, inventory control and production schedule. The data available at one place could be used centrally for the retrieval of required information and information generated by one department could be shared for others to prepare the reports.

The following table depicts the data availability, tools employed and output expected in marketing area.

Table 7.2: Data Availability, Tools Employed, and Output Expected in Marketing Area

Data available	Tools employed	Output
Sales	Forecasting Methods	Expected Production
Advertisement Expenditure	Optimization Models	Profit Maximization
Territorial Data	Methods of Market Penetration	Market Expansion
Consumer Preference	Brand Switching Tools	Market Segment
Price	Comparative Analysis	Pricing Strategy
Movement of Goods	Inventory Tools	Minimizing the Storage Cost

The development of marketing information system should consist of various sub-systems on marketing research, marketing planning, sales analysis and market control. The inputs, processing, and outputs for these sub-systems are as follows:

Marketing Research

Inputs

- Based on external data collated from the industry primarily constituting sales performance analysis of a variety of brands in various market segments of various companies
- Industrial infrastructure, marketing, production strength and weakness
- Industrial advertising drive and impact
- Industrial pricing policy and consequences
- Internal data compiled in the company constituting sales performance analysis of various brands of the company in various market segments;

Information Systems

- Strategic marketing attribute and forecast data versus actual sales performance data
- Advertising budget and actual spending
- Infrastructure budgets
- Prices of products and sales figures

Outputs

- Intensive reports on industry wide sales performances
- Inferences of results compared to forecasts with respect to past marketing strategy, advertising and pricing policies
- Review of marketing strategy, advertising and pricing policies
- Market share of the company

Marketing Planning

Inputs

- Summarized data from market research

Outputs

- Review of sales forecasts
- Product mix strategic reports
- Market mix strategic reports

Sales analysis

Inputs

- Sales data by region/product/personnel in terms of value and quantity
- Projected sales data by region/product/personnel for comparison with the actual sales
- Sales data customer wise with the transaction data for invoicing system

Outputs

- Sales reports region/product/personnel in terms of value and quantity
- Variance analysis on actual sales versus projected sales

Marketing Control

Inputs

- Sales data of various products of a company
- Marketing expense data
- Summarized output sales analysis constituting sales performance data personnel/region/product wise
- Sales transaction data from invoicing and sales systems

Outputs

- Variance analysis
- Age analysis
- Tactical reports on desirable and undesirable trends

The typical transaction in the marketing sub-system is depicted through the Figure 7.6 :

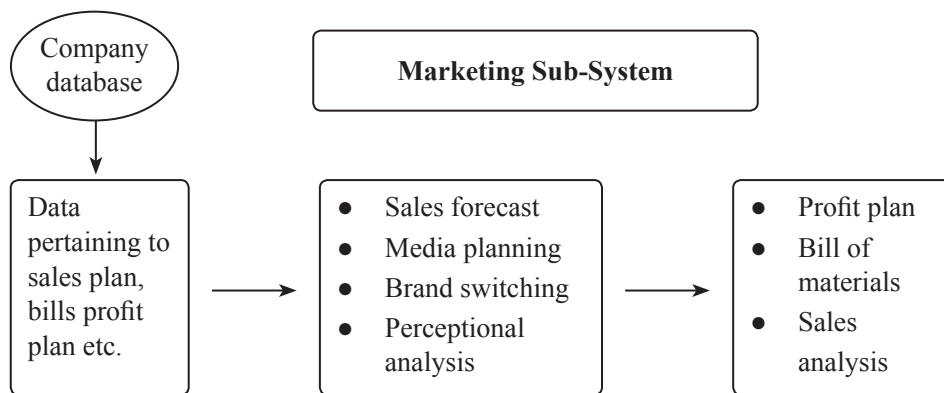


Figure 7.6: Typical Transaction in the Marketing Sub-System

To integrate the activities of marketing, sub-systems are broadly classified on the following types.

- 1) Control systems
- 2) Planning systems
- 3) Market research systems
- 4) Sales systems

Activity A

Prepare a brief report on the market planning and sales analysis of a company of your choice. What are the benefits derived out of computers and MIS.

.....

.....

.....

.....

7.8.2 Financial Management System

Financial accounting deals with collection, recording, and evaluation of financial data. Business enterprise requires systematic maintenance of their records that help for preparing the financial statements like Profit & Loss Accounts and Balance sheet. Accounting is considered as an information system because it has inputs of financial data, processing by evaluation and outputs through financial reports. The importance of financial information system for the different users in the managerial area can be depicted as:

Nature of information	Ultimate user
Cost planning and cost control of operations Profitability of the firm Strategic and tactical decisions	Management
Profitability of the firm Soundness of the invest Growth prospects of the firm	Shareholders & investors

Liquidity of the firm Profitability and financial soundness	Creditors
Statement of salaries, wages and bonus	Employees
Managing the industrial economy of the Country, Collection of sales, excise and Other taxes	Government
Financial growth of the firm Social role of the firm in different sectors	Consumers & public

To fulfill the needs of financial information the information system used in most of the companies may be classified as under:

- 1) Financial accounting systems - To provide information to Government, investors, and other parties in the form of reports
- 2) Management accounting system - To provide reports to managers for strategic and tactic decisions
- 3) Cost accounting system – To provide information about cost planning and control operations to the managers.

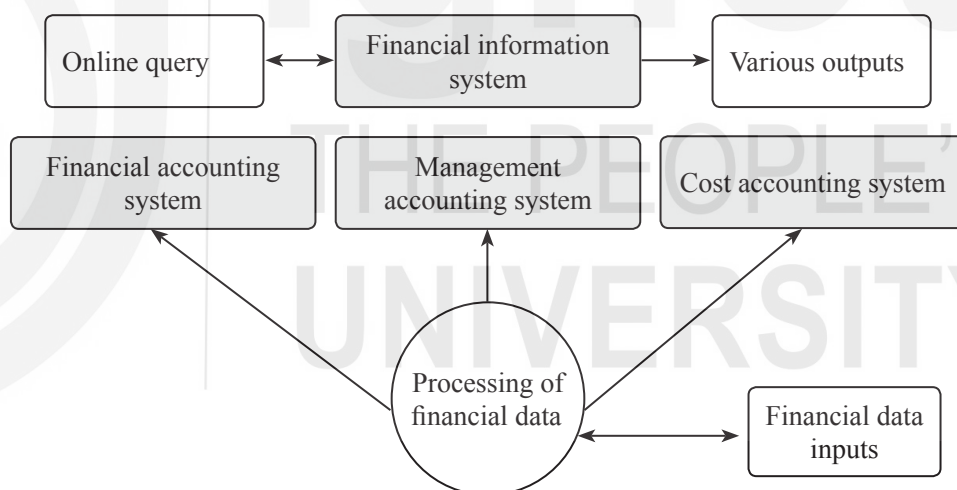


Figure 7.7: Financial Information system

The objectives of implementing computerized financial accounting system include:

- Maintaining account books
- Preparation of general ledger
- Generating accounts receivable and accounts payable statements
- Generating profit & loss account and balance sheet
- Generating updated financial data for other systems

Inputs to the system

The input data from the financial system are from

- Cash vouchers

- Bank vouchers
- Sale vouchers
- Purchase vouchers
- Journal vouchers

Outputs to the system

- The outputs expected from the system are:
- Account books like cash book, bank book sale book etc
- Trial balance
- Trading account
- Profit & Loss account
- Balance sheet
- Accounts receivable statement
- Account payable statement

The Financial management is the most important aspect in managerial functions. Managers quite often require the financial position of the organization for various reasons like planning, monitoring, performance evaluation etc of the companies to decide about further action. In traditional method of information retrieval different kinds of books, ledgers and financial records have been kept to follow up the financial position and by end of each month status used to be calculated. Computerization of financial record not only simplifies the procedure but also is more accurate. With computerization on-line information could be accessed. As most of the companies have gone for the computerized data record, creating a database with required entries will facilitate the storage of information and any kind of manipulation based on the database. A few advanced applications in financial transactions are e-commerce, e-business and on-line banking.

System development for financial management is not only to make the computerized billing or providing the receipts for the transactions. The system integrates number of financial transactions of the organization.

7.8.3 Human Resource Management

The manpower utilized in the organization is treated as a valued asset. The management of this asset is just like management of any other source in an organization. Organizational, effectiveness and efficiency, growth of the business, sustainability for the competitiveness can be seen through proper management of human resources. HRD management emphasizes an optimum utilization of human resources by formulating consistent and coherent policies aimed at promoting commitment to the organization. The information system on HRD is restricted to personnel management systems whose role is to assist in

- Recruitment,
- Placement,
- Training and development,
- Compensation and
- Maintenance

The vital area covering the above activities is called Human Resource Planning. HRD system binds functions of personnel management systems with other functions of organization such as marketing, finance, production, and inventory.

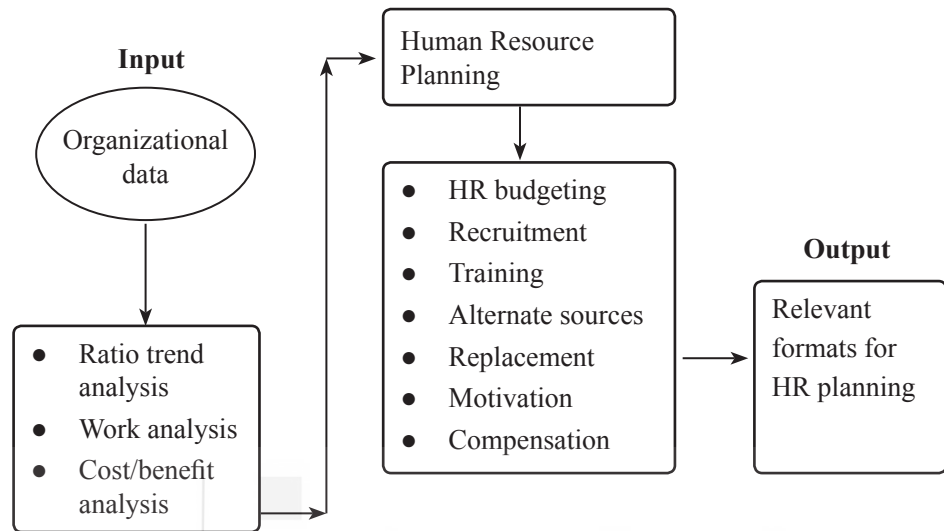


Figure 7.8: HR Information system

7.9 TRANSACTION PROCESSING SYSTEM

Management information system, in general, consists of all business transactions of the organization. Every information system has many subsystems like transaction processing system, decision support system, office automation system etc, to have a larger utility of database created for business analysis. The decision makers do retrieval of information from the master data through a set of well-defined tools. For example, to calculate the net salary of an employee, we need to supply different inputs like basic salary, number of days present, existing DA, and HRA levels. Accounts department uses this information to calculate the final receivable salary of an employee. This process is called transaction process of the data for payroll calculation. As this is a part of the information system, it can be called as a subsystem. Preparation of transaction documents such as pay slips, invoices, and receipts is important and routine task performed in managerial functions. The general view of transaction processing system is shown in Figure 7.9:

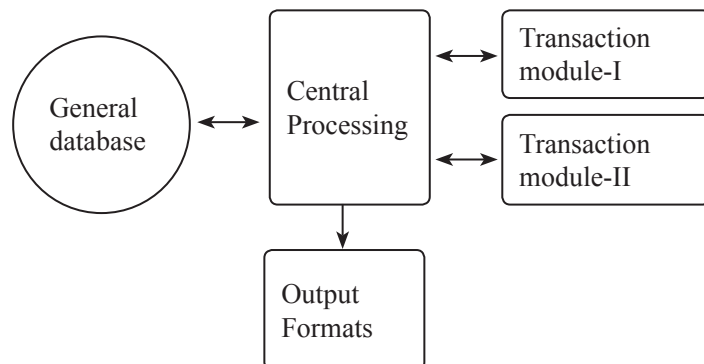


Figure 7.9: General view of Transaction Processing System

Data manipulation to get the desired information through the specific tools of management is called transaction process. As applicable in other cases, different areas of management like HRM, Finance, Operations, and Marketing have various business transaction processing systems and hence the sub systems to generate required output are needed. We will discuss some examples of such sub systems for transaction processing.

7.9.1 Human Resource Management Sub-System

In Human Resource Management, the Personnel Manager has the responsibility of executing multifarious activities pertaining to recruitment, training of the employees, retirement benefits of the employees etc. Planning, executing, and controlling of these activities about the existing employees is carried out by the personnel manager through the information system called personnel information system developed for this purpose. The personnel information system deals with the flow of information about people working in the organization as well as future personnel needs. In most of the organizations, the system is concerned primarily with five basic subsystems of the personnel function: recruitment, placement, training, compensation, and maintenance.

Let us discuss three of these subsystems separately:

1. Recruitment
2. Retention/service conditions
3. Retirement

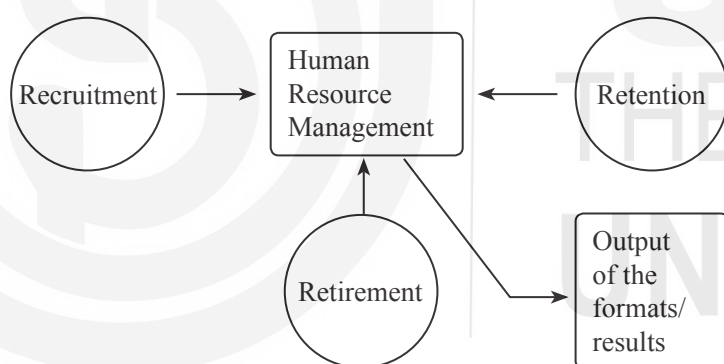


Figure 7.10: Transaction Modules in Human Resource Management

Recruitment

Role of personnel department starts with recruitment of personnel. If properly managed, the recruitment system forecasts personnel needs and skills and recruits the personnel at the proper time to meet organizational needs. Personnel department on the basis of retirements/resignations either identifies requirements, or concerned department sends requirement to personnel department. On the basis of the requirements, personnel department initiates process of recruitment, which is divided into the following steps:

- Finalizing the requirement for various posts
- Publicity as per requirement
- Receipt of application forms from candidates

- Initial screening
- Interviews
- Final selection
- Issuing appointment letters

Development of proper information system for generation of reports pertaining to the above cases, by giving the necessary inputs, lead to a Sub-System for this purpose.

Retention/Service conditions

This is an important aspect in the functioning of personnel department. Personnel department maintains records of all the employees about the nature of the job, department etc. The major activities undertaken and related data requirements here are:

- Training/Seminars/Refresher courses for employees
- Job
- Promotion
- Increments
- Leave accounting
- Performance reports
- Salary/compensation
- Organizational structure
- Job description
- Liaison

Information on each of the areas is compiled from the master database created and made available in the organization.

Retirement

Personnel department keeps track of retiring employees. It completes all the formalities necessary and issues orders for providing retirement package through account department. The calculation requires various inputs about the employees from the personnel records.

There are other subsystems also like Compensation & Maintenance etc. Another interesting subsystem could be Skills inventory program, which identifies and locate the talent resources of the organization to maximize their use. The concept of “Skill banks” or “manpower assessment program” is applied. This could be one of the subsystems to develop and support the main system.

7.9.2 Marketing Management Sub-System

As we have seen in the previous unit, the business function of marketing deals with different activities like planning, promotion, and sale of existing products and exploring the market trend for the existing and new products. Information system specially designed taking into account all these factors can only supply the required competitive information for the benefit of customers and management.

Different categorical decision systems like strategic, tactical and operational information help in the activities of the marketing managers in product planning, pricing of the product, planning for advertising, sales promotion & expenditure, forecasting market potential for new and existing product and deciding about the distribution channel etc.

Control reporting systems support the efforts of marketing managers to control the efficiency and effectiveness of selling & distribution of products and services. Analytical reports provide information on firm's actual performance versus planned marketing objectives.

The major marketing subsystems are:

- Sales Management
- Sales forecasting
- Advertising and promotion
- Marketing research



Figure 7.11: Transaction modules in marketing area

Sales Management

Sales Management system provides information to help sales managers plan and monitor the performance of the sales organization. In most firms, this system produces sales analysis reports, which provide information analysis of

- a) Product sales trends
- b) Product profitability on a product-by-product basis
- c) The performance of each sales region and sales branch
- d) Sales persons performance

To fully support the sales systems, invoices should contain information about product type, product quantity, price discount terms, customer identity & type, sales region and sales person.

The sales reports must contain information about the profitability of products, product lines, sales territories, and individual salespersons. Profitability reporting requires information about product administration and selling costs.

The sales analysis system also provides information used to evaluate the performance of sales personnel. By reclassifying sales by sales person rather than by product or sales unit, it is easy for the sales supervisors to identify whether each salesperson has met the assigned sales quota. If marginal costs are also assigned by salesperson, each salesperson's contribution margin can be established as one criterion for performance evaluation.

The primary types of information as inputs used for the sales analysis subsystem are:

- i) Information about customer invoices and back orders that is provided by the sales transaction processing systems
- ii) Information derived from the oral and written reports of sales personnel
- iii) Product cost information obtained from the cost accounting information systems
- iv) Information about the sales transactions may be available from the sales files or the customer files.

Sales Forecasting

Sales forecasting is important for planning sales and promotion strategies and for post period evaluation of salesperson's performance. Sales forecasts are generally the first step in the profit planning cycle of the entire organization e.g. they initiate the short-term financial and marketing planning cycles. If forecasts for the period are seriously in error, the organizations entire budgeting activity will be based upon the wrong level of activity.

The basic functions of sales forecasting can be grouped into the two categories of short-range forecasting and long-range forecasting. Short-range forecasting deals with forecasts of sales for periods up to one year, whereas long-range forecasting is concerned with sales forecasts for a year or more into the future. Forecasting requires the application of mathematical and statistical methods, which consists of structured procedure. Methods like time series analysis, regression models etc., are used for sales forecasting.

Production planning, work force planning, and production scheduling are based on sales forecasting. Materials are ordered and personnel are employed and reassigned according to the expected sales pattern for the next period. If sales are overestimated either finished product inventory levels or materials inventory will be excessive, or personnel will have to be laid off. Conversely, if the sales forecast is too low, either sales would be lost because of inventory shortages, or higher costs will be incurred as a result of overtime and need to secure additional materials in a hurry.

The basic information needed for sales forecasting is as follows:

- i) Past sales, provided by the sales analysis information system
- ii) Market conditions, provided by the market information system
- iii) Competitor's activities provided by the competitor intelligence information system
- iv) Promotion and advertising plans, provided by the information system for this function

In similar way, sub-systems for other application systems like Finance Management, Production and Operations Management, Inventory Management etc. exist for corresponding functionalities.

Activity B

Consider an organization dealing with on line sales by maintaining inventory of items. There are suppliers who supplies items to the organization for sales. Outline the various functional activities for the transactions involved.

.....

7.10 DSS AND EIS

DSS stands for Decision Support System. It is software which supports an organization in taking decision. Executive Information System (EIS) is a data driven DSS that is intended to provide business intelligence to middle and senior level managers. In both the systems decision making process is involved. Let us now understand the decision making process first of all.

7.10.1 Decision Making Process

A **decision** is basically a resource allocation process that is irreversible, except that a fresh decision may reverse it or overrule the earlier one. The **decision maker**, having authority over the resources being allocated, makes a decision. He makes the decision in order to further some **objective**, which is what he hopes to achieve by allocating the resources. The decision might not succeed in achieving the objective. One might spend the funds and yet, for any number of reasons, achieve no acceleration at all. For example: To accelerate an R&D program is an objective, not a decision. To allocate the funds in an effort to accelerate the program is a decision.

Simple decision is one in which there is only one decision to be made, even having many alternatives.

A decision may be **goal** oriented for some degree of satisfaction for a given objective. Objective may be driven by a decision but goal is always target/result oriented. A decision may employ **decision analysis**; a structured thought process to attain desired results. In doing this, we can distinguish three features of the situation: **alternatives, uncertainties and outcomes**. Decision analysis thus constructs **models**, logical or even mathematical, representing the relationships within and between the features of situation. The models then allow the decision maker to estimate the possible implications of each course of action that he might take, so that he can better understand the relationship between his actions and his objectives. Someone who buys a lottery ticket and wins the lottery obtains a good outcome. Yet, the decision to buy the lottery ticket may or may not have been a good decision.

Decision maker may adopt **strategy**, which is a collection of actions. The outcome of these actions may be uncertain, but the possible outcome is attainment of that goal set by the decision maker. It is necessary to mention

about the **risk** at this juncture that a decision maker often thinks while taking the decisions. Risk can be defined as the possibility of an undesirable result. The risk is linked with monetary benefits or loss and thus, it can be governed by **risk tolerance**, an individual's attitude toward decision and the risk involved.

With the above background, we can now distinguish between **strategic and tactical decisions**. Strategic decision affects entire organization or major part of it vis-à-vis organizational objectives and the policies. It has long lasting effect on the organizational system and generally taken at the highest management level. On the other hand, tactical decision or management control decision affects a part of the organization for a restricted or short time. The tactical decision takes place within the context of existing strategic decisions. Thus, the contextual and effective management depends on good and timely information. The decision making process may use various techniques depending on the situation. We can define decision-making as an activity of deciding appropriate action in particular situations. With relevant and useful information the decision-making may reduce the uncertainty.

We have studied earlier that Simon provided a general model of any decision making process. The Simon model describes the sequence of decision making as: Intelligence, Design, Choice, and Implementation.

7.10.2 Decision Support System

With the understanding of decision making process DSS can be defined as the system that supports the decision making. There are various definitions of DSS that links to information retrieval with the help of computers, to use of information, to support managers. Let us see that how some of these definitions describe DSS:

Table 7.3: Various definitions of DSS

S. No.	Definition given by	Definition
1.	G.W. Peter Keen and Michael S. Morton	Decision support implies the use of computers to a) assist managers in their decision processes in semi-structured tasks; b) support rather than replace, managerial judgment; c) improve the effectiveness of decision making rather than its efficiency.
2.	Ralph H. Sprague	Information system featuring an integrated system composed of decision models, database and decision maker to support decision-making.
3.	James Hick	An integrated set of computer tools that allow a decision maker to interact directly with computers to create information useful in making unanticipated semi-structured and unstructured decisions.
4.	Carroll W. Frenzel	Decision Support System are analytical models used to increase managerial or professional decision making by bringing important data to view.
5.	William E. Leigh and Michael E. Doherty	A set of computer based tools used by managers in connection with his or her problem solving and decision making duties.

These definitions points that :

- 1) decision support systems are information systems;

- 2) decision support systems are used by managers;
- 3) decision support systems are used in making decisions;
- 4) decision support systems are used to support, not to replace, people;
- 5) decision support systems are used when the decision is semi-structured or unstructured;
- 6) decision support systems incorporate a database of some sort; and
- 7) decision support systems incorporate models.

One word that attracts us is “Model”. Let us see the following figure that describes a simple DSS model :

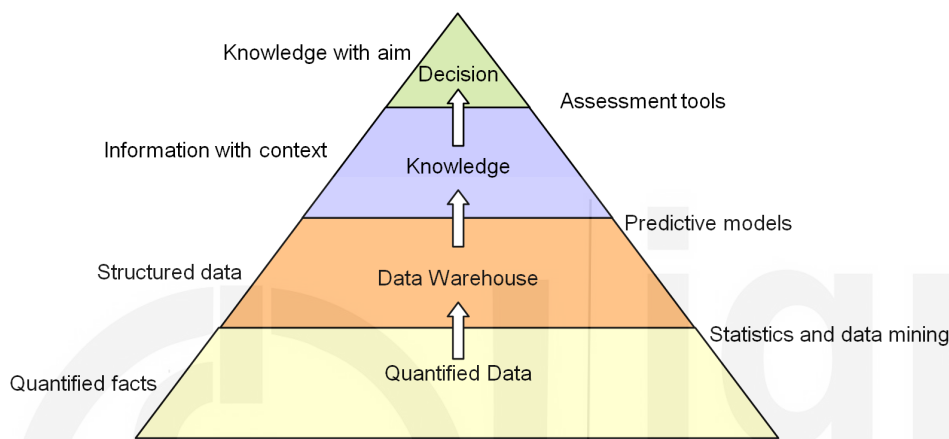


Fig. 7.12 : DSS Model

As can be seen, the decision knowledge emerges from the quantified data using statistical and data mining tools, predictive models and the assessment tools. The qualified data means that the data used to derive quality information must be relevant, accurate, complete, authenticated, up-to-date, in-time available, and easily accessible. Data warehouse is basically repository of organizational data. You will learn about data warehousing and data mining in other Unit ahead.

We also come across Group DSS for taking group decisions using decision support system. GDSS are interactive computer-based systems that facilitate decision-makers working together as a group to arrive at a solution for unstructured problem. The GDSS provides mechanisms to help the users to coordinate and keep track of on-going projects, allow them to work together thru computer-supported communication, collaboration, and coordination. Typical applications of GDSS include email, awareness and notification systems, videoconferencing, chat systems, multi-player games, and negotiation systems. The software developed for GDSS focuses principally on assisting brainstorming and mechanizing voting, two of the rare events in business meetings. Group support systems are designed to support group decision-making through specialized software, hardware and decision support tools. This can be defined as a combination of computer, communications and decision technologies working in tandem to provide support for problem identification, formulation and solution generation during group meetings. Broadly, the fundamental goal of GDSS is to support the exchange of ideas, opinions, and preferences within the group.

GDSS are believed to improve the quality of group decisions by minimizing process losses and maximizing **process gains**. Process gains occur when certain aspects of the meeting improve the eventual outcome or result and process losses delay or reduce the final outcome. Thus, the overall meeting outcome is reliant upon the process gains versus the process losses.

Among other terms used in GDSS like Groupthink, Communication Media Richness, Groupware there is a term **War room** used in GDSS. It is basically the electronic meeting room or electronic boardroom, created for GDSS by networking computers for each member participant. It is equipped with a large screen video display that can be seen by all the participants. There is a facilitator to control the display and have access to all member's computers. The software used have the capabilities to plan a session and group members can propose agenda items through it. It can organize and structure member's comments and can record them as organisational memory for longer and effective use. These days, we come across many such tools available for group meetings apart from those available in social media.

To build a DSS in organization, one has to consider the spectrum of DSS that organization is going to use. The following are to be well thought-out to approach systematically :

- Strategic, tactical and operational decisions of the organization,
- Unstructured, semi structured and structured decisions,
- All levels of management and knowledge workers in the organization,
- Major functions, product or line of business,
- Hierarchy in operations, and
- Geographical division of the organization.

7.10.3 Executive Information System

An **Executive Information System (EIS)** is a support system for executives of an organization. It is a kind of decision support system used in organizations to help executives in decision making. It does so by providing easy access to important data needed in an organization to achieve strategic goals. It is used for monitoring organization's performance and also for identifying opportunities and problems.

Today every performing organization has its own website and connectivity in terms of intranet and also through internet integrating all of its branches at distant locations. Employees can access company data to help make decisions in their workplaces, departments, divisions, etc. This enables employees to provide relevant information and ideas, advises and suggestions for betterment of the organization.

EIS as a support system is intended to be used directly by senior managers to support unscheduled strategic management decisions. Often such information is external, unstructured and even uncertain. The key characteristics of EIS are :

- Detailed data – EIS provides absolute data from its existing database.
- Integrate external and internal data – EIS integrates external and internal data. The external data collected from various sources.

- Presenting information – EIS represents available data in graphical form which helps to analyze it easily.
- Trend analysis – EIS helps executives of the organizations to data prediction based on trend data.
- Easy to use – It is a very simplest system to use.

7.10.4 Difference between MIS, DSS, and EIS

With the exposure to all the above information systems, let us find out the differences between DSS, MIS, and EIS. Table 18.2 enlists some basic differences between DSS, MIS and EIS. As the name implies, the later two are the systems that provide information that may or may not be used for making a decision whereas the support information provided for deciding on the policy, planning or implementation is the basic component of DSS. Let us find out the characteristics of the three systems:

DSS

- DSS generally provide support for unstructured, or semi-structured decisions (decisions that cannot be described in detail)
- DSS problems are often characterized by incomplete or uncertain knowledge, or the use of qualitative data
- DSS will often include modeling tools in them, where various alternative scenarios can be modeled and compared
- Investment decisions are an examples of those that might be supported by DSS

MIS

- MIS is generally more sophisticated reporting systems built on existing transaction processing systems
- Often used to support structured decision making (decisions that can be described in detail before the decision is made)
- Typically will also support tactical level management, but sometimes are used at other levels
- Examples of structured decisions supported by MIS might include deciding on stock levels or the pricing of products

EIS

- EIS supports a range of decision making, but more often than not, this tends to be unstructured
- EIS supports the executive level of management, often used to formulate high level strategic decisions impacting on the direction of the organization
- These systems will usually have the ability to extract summary data from internal systems, along with external data that provides intelligence on the environment of the organization
- Generally these systems work by providing a user friendly interface into other systems, both internal and external to the organization

Let us now explore differences among the three information systems based on the dimensions (Table 9.4).

Table 7.4: Difference between DSS, MIS and EIS

Dimension	DSS	MIS	EIS
Focus	Analysis, decision support	Information processing	Status access
Typical users served	Analysts, professions, managers (via intermediaries)	Middle, lower levels, sometime senior executives	Senior executives
Impetus	Effectiveness	Efficiency	Expediency
Application	Diversified areas where managerial decisions are made	Production control, sales forecasts, financial analysis, human resource management	Environmental scanning, performance evaluation, identifying problems and opportunities
Database(s)	Special	Corporate	Special
Decision support capabilities	Supports semi-structured and unstructured decision making; mainly ad-hoc, but sometimes repetitive decisions	Direct or indirect support, mainly structured routine problems, using standard operations, research and other models	Indirect support, mainly high level and unstructured decisions and policies
Type of information	Information to support specific situations	Scheduled and demand reports; structured flow, exception reporting mainly internal operations	News items, external information on customers, competitors and the environment
Principal use	Planning, organizing, staffing and control	Control	Tracking and control
Adaptability to individual user	Permits individual judgment, what-if capabilities, some choice of dialogue style	Usually none, standardized	Tailored to the decision making style of each individual executive, offers several options of outputs
Graphics	Integrated part of many DSS	Desirable	A must
User friendliness	A must where no intermediaries are used	Desirable	A must
Treatment of information	Information provided by the EIS/or MIS is used as an input to the DSS	Information is provided to a diversified group of users who then manipulate it or summarize it as needed	Filters and compresses the information, tracks critical data and information
Supporting detailed information	Can be programmed into DSS	Inflexibility of reports, cannot get the supporting details quickly	Instant access to the supporting details of any summary
Model base	The core of the DSS	Standard models are available but are not managed	Can be added, usually not included or limited in nature

Construction	By users, either alone or with specialists from IS or IC departments	By vendors or IS specialists	By vendors or IS specialists
Hardware	Mainframes, micros or distributed	Mainframes, micros or distributed	Distributed system
Dimension	DSS	MIS	EIS
Nature of computing packages	Large computational capabilities, modeling languages and simulation, applications and DSS generators	Application oriented, performance reports, strong reporting capabilities, standard statistical, financial, accounting and management science models	Interactive, easy to access multiple databases, on-line access, sophisticated DBMS capabilities and complex linkages

Adapted from the Information Systems resource materials by Denis Manley, School of Computing at Dublin Institute of Technology.

Activity C

- Let us take an example of a Travel Agent selecting destinations for Clients based on the Client's expressed interest in vacation activities and the Agent's knowledge of what is available in various locations. Do you think that a computerized DSS will help in decision-making? If not, why? If yes, how?

.....

.....

.....

7.11 SUMMARY

In this unit, an attempt has been made to explain the importance of information, information generation as a system and role of information in various areas of functional management. We have discussed some of the key points (environmental and competitive) to be kept in mind while going for the development of information systems.

All organizations have similar needs to build up from the basic inputs a good quality database each and then to evaluate selective data, analyzed data, inter-linked data and aggregated data to derive suitable MIS support to decision making at the corporate level. We have discussed various functional management systems to illustrate with input, output and processing aspects thus generating the organizational database to help taking business decisions by the managers. Transaction processing was discussed as the sub system of functional management system. EIS and DSS have been discussed with tabulating the difference of these with MIS. Also Group DSS has been discussed for arriving at decisions with involvement of all concerned.

7.12 UNIT END EXERCISES

- Discuss the importance of information system for business decisions.
- Distinguish between data, information and knowledge.

3. Visit some of the organizations to identify the information systems used for better management. Suggest the ways and means for the improvement.
4. Identify and name some of the computerized information systems used in production management, HRD and marketing.
5. Identify the role of transaction processing system in the development information system.
6. List the analytical methods, which could be used effectively in the modules of human resource management for the development of subsystem.
7. How does a computerized Personnel Information system help in managing human resources more efficiently?
8. Why do both EIS and DSS are required in decision making process? Illustrate with two examples.
9. What is the role of Group DSS in arriving at decisions? Name two such software which facilitate group meetings over internet.
10. What is the role of social media in supporting decision making process in an organization?

7.13 REFERENCES AND SUGGESTED FURTHER READING

Adelman, Leonard, *Evaluating Decision Support and Expert Systems*, New York : John Wiley and Sons.

Clifton, H. D., *Business Data Systems*, Prentice-Hall: Englewood-Cliffs.

Mallach, Efrem G., *Decision Support and Data Warehouse System*, New Delhi :Tata Mcgraw Hill.

McGrath, Joesph Edward (1984) *Groups : Interaction and Performance*, New Jersey : Prentice Hall.

Murdick, Robert G., *MIS: Concepts and Design*. Englewood Cliffs, N.J. Prentice-Hall.

Murdick, and Thomas C. Fuller., "Subsystems for MIS". *Journal of Systems Management*, June 1979.

Murthy CSR, *Management Information Systems: Text and Applications*, Third Edition, Himalaya Publishing House, Mumbai

Robert G. Murdick, Joel E. Ross and James R Claggett. *Information systems for Modern Management*, Prentice Hall of India.

Sadgopan S., *Management Information Systems*, Prentice-Hall if India Private Limited, New Delhi, India

Walker D., *Computer Based Information Systems*, Elsevier Science Pub Co; 2nd edition.

<https://www.includehelp.com/MIS/executive-information-system-eis.aspx>
site accessed on 27th November 2012

UNIT 8: INTEGRATED APPLICATIONS

Structure

- 8.1 Introduction
- 8.2 Objectives
- 8.3 Integrated Software Applications and Their Importance
- 8.4 Advantages and Business utility
- 8.5 Implementing Integrated Applications
- 8.6 The Competition
- 8.7 Business Software Solutions from Oracle
- 8.8 Business Software Solutions from SAP
- 8.9 Business Solutions from Microsoft
- 8.10 Summary
- 8.11 Unit End Exercises
- 8.12 Reference & Suggested Further Readings

8.1 INTRODUCTION

Integrated application software systems have become necessary for organizations to compete and do the business in controlled and profitable manner. They need to choose a system that suits their business processes. Integrated systems are the systems that help in integrating business processes and management functions. They provide a transparent and integrated view of all the activities that are going on in the business. Unlike earlier days, when data used to be collected from different service divisions/nodes and then compiled in the form of a summary report, presently the software does the compilation and the integration work. This facilitate in maintaining consistency of data by making same data available to all concerned in the organization. It also minimizes the time taken in accomplishing the work. There are many software packages that integrate the activities of different service divisions and respective processes. This way they facilitate businesses to devote more time on productivity. However there are different needs of different business and one should be careful in deciding about a software system that would be useful to the organisation. Any company looks for these packages to integrate its corporate functions like finance, manufacturing, customer relations, human resources and so on. This is a critical activity for organisations. As a successful manager you should be able to understand the system requirement of your company. You should be able to define the information flow, information requirements and information usage so that you gain competitive edge. This unit is thus an attempt to keep you updated on the latest in the business solutions area.

8.2 OBJECTIVES

After reading this unit you should be able to

- Explain the concept of integrated software applications;

- Describe their advantages, business utility, and implementation issues; and
- Discuss the business solutions offered by market leaders.

8.3 INTEGRATED SOFTWARE APPLICATIONS

Literal meaning of “Integration” is combination or amalgamation. In terms of computer terminology, “Integration” is a broad term for any software that serves to join together separate and usually already existing applications. Integrated software applications for business give you the ability to integrate the diverse information sources pertaining to your business holistically into a single framework. This integrated information can then be shared by applications such as Customer Relationship Management (CRM), Supply Chain Management (SCM), Business Process Re-engineering (BPR) systems as so on. The amalgamation of integrated software applications is called Enterprise Resource Planning, widely known as ERP.

Now, let us see these terms with regard to their respective definitions with little more about them:

ERP: Short for *Enterprise Resource Planning* is a business process management system that integrates all facets of the business, including planning, manufacturing, sales, marketing, and front-end services like human resource management and finance management covering accounting, budgeting, deposits etc. The term ERP was coined by the research firm Gartner around 1990. As the ERP methodology has become more popular, software applications have emerged to help business managers implement ERP in business activities such as inventory control, order tracking, customer service, finance and human resources. Integrated software applications are available now for almost all business houses automating many back office functions related to technology, services and human resources. They can also be widely seen today in use for various services in Hospitals, Banking, Education and many more. ERP systems are now vast repositories of information helping organizations to generate reports that can spotlight the performance of every aspect of the business. There are countless applications available, designed for different industries, business models and challenges, and ERP acts as command central for a vast network of application software.

CRM: Short for *Customer Relationship Management* entails all aspects of interaction a company has with its customer, whether it is sales or service related. Computerization has changed the way companies are approaching their CRM strategies because it has also changed consumer-buying behavior. With each new advance in technology, especially the proliferation of self-service channels like the Web, Bluetooth and WAP (Wireless Application Protocol) phones, more of the relationship is being managed electronically. Organizations are therefore looking for ways to personalize online experiences (a process also referred to as mass customization) through tools such as help-desk software, e-mail organizers and Web development applications. Today we find sales and customer services related communications very actively and promptly disseminated through emails, messaging services, social media giving the comfort to consumers

of getting updated in no time. Once we place an order for purchase of some items through online we get instantly updated about status of order and other related business activities such as payments confirmation etc.

SCM: Short for *Supply Chain Management*, the control of the supply chain as a process from supplier to manufacturer to wholesaler to retailer to consumer. Supply chain management does not involve only the movement of a physical product (such as a microchip) through the chain but also, like CRM, any data that goes along with the product (such as order status information, payment schedules, and ownership titles) and the actual entities that handle the product from stage to stage of the supply chain.

There are essentially three goals of SCM: to reduce inventory, to increase the speed of transactions with real-time data exchange, and to increase revenue by satisfying customer demands more efficiently maintaining unified communications at all levels giving the present status to the customers. It is used in reference to software applications that enable more efficient management of the supply chain.

BPR: Short for *Business Process Re-engineering* is aimed to make radical changes in an organization from the ground up in an aim to improve performance and make more efficient use of resources. It focuses on the analysis and design of workflows and business processes which are nothing but the steps and procedures that govern how resources are used in the organization to meet the needs of customers or markets. It helps organizations to fundamentally rethink how they do their work in order to improve customer service, cut operational costs, and compete globally. During the process of re-engineering it uses IT to organize data, doing data analysis, and maintain project trends etc. for success of the business. Using the technological innovations organizations going for re-engineering replace old methods of business operations with cost-saving innovative technologies by using integrated software applications to radically transform business operations. Re-engineering focuses on re-designing the process as a whole in order to achieve the greatest possible benefits to the organization and their customers.

Importance and use of software applications for various services in organizations is well known today. We hardly find any sector today untouched by computers. Be it education, factory, agriculture, banking, hospitals, Research & Developments, Government offices everywhere we find use of computers for the daily business and various services. Now companies are giving high importance to software integration without which it is almost impossible to compete in the market. Today nobody wants to access data stored in multiple systems. They want to build strong links between business systems and make information flow better. Let us take an example, suppose you want to place an order online using your smart phone or laptop or desktop computer. For connectivity you have to have, of course, the internet. An integrated software solution, on one hand, takes that order, shifts it and allocates them to the manufacturing plant and on the other hand places order for the raw materials on the basis of the stock, updates the financial position of the company with respect to suppliers and the inventory and so on. Many people have given different names to the

integration of ERP, SCM, BPR and CRM. These names include online business or e-business, c-business (collaborative e-business), m-business or mobile business etc.

There are many software that do these integration activities. At present, in the year 2021, the top ten software systems for the business process integration are SAP Business One, NetSuite, SAP, Sage Intacct, SAP S/4HANA, Acumatica, Tally, Odoo ERP, SAP ByD, and Oracle EBS (Source: <https://www.g2.com/categories/erp-systems>, 9th August 2021). Rating is based on survey taken by the company primarily with the user satisfaction level and the popularity. Microsoft Dynamics NAV is another such system by Microsoft for process integration. It is worth mentioning here that Odoo ERP is Open Source Software meaning thereby that software codes are available to users for using almost free. Users have to further modify them as per the specific requirements. Unlike Proprietary software, it does not require user licenses to use the software. But it does require in depth knowledge of software coding to understand and modify to suit the user requirements. There are many such open source software in the market. To name few of them Adaxa Suite, ERP5, ERPNext, inoERP, LedgerSMB, SQL-Ledger, WebVella etc are available globally. With the advancements in IT, all these services are also available with just a click on web link with computer servers and storage systems. Users do not have to worry about space, neither for electricity or air-conditioning for physically keeping servers and storage systems at their premises. All are available in Cloud i.e. physically at other IT service provider location.

8.4 ADVANTAGES AND BUSINESS UTILITY

Integrated software packages help in real business growth. Seeing the wide spread availability of such solutions from IT solution providers, companies today put on priority to integrate their diverse business processes to simplify operations for faster decision-making. Companies realise that if they have to survive and grow, they have to use tools that can provide quicker and useful information and cut costs to increase efficiency. IT has given businesses a chance to generate information in real-time and thus grab opportunities that were non-existent or unseen earlier. With the development of cutting-edge technology, we have seen evolution of many systems that has adapted market demands. IT is readily providing business solutions now. Even complete project undertaken by a company can be very well managed and monitored today using Project Management Systems (PMS). They are designed to assist project managers and their teams organize the step-by-step process of a project, identify priorities, and monitor project tasks, costs, contacts, deadlines and staff. Complex projects that require staff, team coordination, budgets and so on can be difficult to manage without the use of project management software. Similarly, Business Intelligence (BI) tools are also available for analytical study for making the business more and more successful.

The integrated software solutions provide a platform to integrate all processes in an organization enabling it to plan, trace and see its 4-m resources (materials, machines, men and money) in the best possible ways to service customers and reduce costs. If you want that your Business should

survive with low overheads and still run efficiently then you have no choice but to opt for IT solutions.

Today, a typical business solution is an integration of ERP, BPR, SCM, CRM, BI, and PMS. While ERP can take care of functions like accounts production planning, payroll and marketing etc., BPR can be used simultaneously to cut-down on all non-value added business processes and paperwork. These systems are the enablers of change in the business for better. With the help of these systems your business can compete better. These systems, combined with the web, help you to reach your customers most cost-effectively. There would be no cost incurred on travel and communication and no cost incurred on setting up offices and employees. The Cloud services available with almost all big players in the market have made the task of maintaining integrated system much more simpler as compared to yesteryears.

E-Business generations are well explained with a pictorial representation in the Figure 8.1 . By this you would be able to appreciate why we are talking about integrations and business solutions and what place it has in the present world. It can be seen that 4th generations business solutions are using web to integrate all systems. ERP is used in a broader term here. In fact the term ERP is the synonym for integration of all the software catering to various business processes in organizations. Worldwide this term i.e. ERP is used now for the total integration of business processes. The software modules, BPR for reengineering the business processes and PMS for project monitoring are used too for the respective tasks. Similarly BI is used as the analytical tool for estimating and forecasting supports for the business based on primary data (at the organization location) and secondary data available through web for the analysis and inferences.

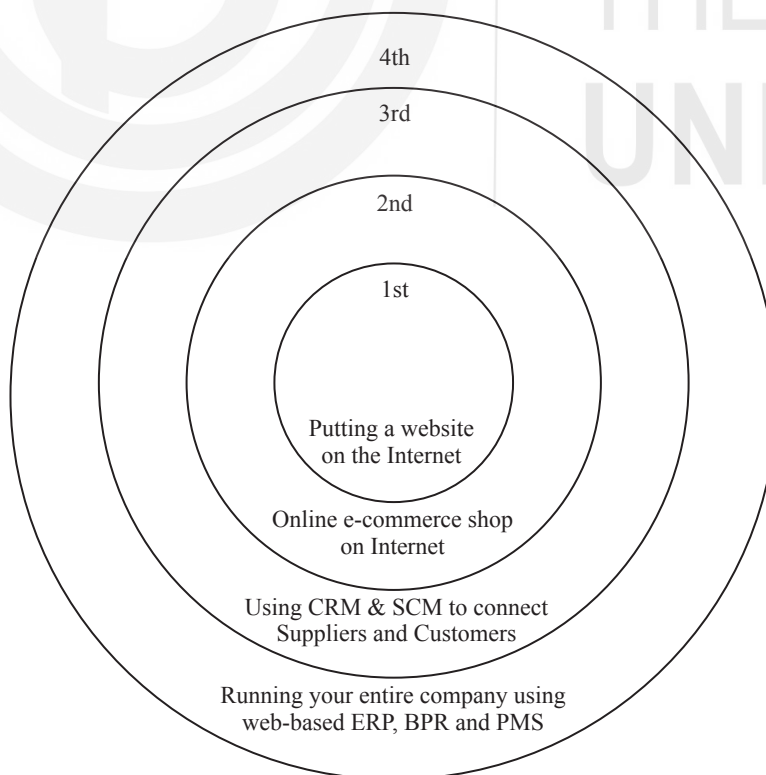


Figure 8.1: Generations of e-Business

Here you should be clear that e-business is similar to e-commerce but it is more than just a simple act of buying and selling services or goods online. In fact, it is the method of utilizing digital information and advanced communication technologies to streamline different business processes – from the initial to implementation phase. E-commerce is just a part of e-business.

8.5 ISSUES IN IMPLEMENTING INTEGRATED APPLICATIONS

The key to establishing a good IT system lies in the proper planning. Businesses should first specify their needs, processes and key data. They should clearly specify the kind of information and its flow. Once these specifications are done, it is the integration of information that is required. Once integration is done, information is accessible to every department around the company. Customers, suppliers and others may also be provided with the facilities to access relevant information with proper access rights, say, status of order, supply, delivery etc.

In most of the cases organizations hire the services of IT service providers. These facilities being from third party require comprehensive study and analysis of business requirements to have the integrated system in place. As-Is and To-Be processes must be well documented for reference now and then by both – the organization and the service provider. As-Is document contains all what is required to be automated and integrated. Process owners available in the various departments of the organizations are the key persons to put forth the existing processes which the organization has been using. This may also contain new requirements to be automated and made part of the new system. ERP and other software packages available for automation are required to be customized to suit the organization. The To-Be document must have mention of all automated and integrated processes which would be delivered to the organization by the IT service provider after completion of the project. As the end result will be the integrated system hence any small mistake at any stage may have great impact on the business and system may halt. Hence all process users must be involved right from giving the requirements and they must be available till the delivery and acceptance. Implementations of new system – automated and integrated applications will never be a success till above is done.

Selecting the right service provider is another issue. Selecting the best and most cost effective solution depends on developing the RFQ/RFP containing the ‘What’ part of the solution requirement comprehensively. RFQ (Request For Quotation) or RFP (Request For Proposal) must contain all requirements that each user department needs in terms of the various outputs/reports once the new integrated software is in place. You will study further about these terms in the Unit on Systems Analysis and Design.

ACTIVITY A

Take the case of your organization or any other organization of your choice. Classify the activities of the organization with respect to the generations of e-business. Mention all the features by virtue of which you have done this classification.

8.6 THE COMPETITION

In this Unit we would introduce to some of the leading players providing IT based business solutions. But before entering what these players are offering, let us see the interesting competition that has been in past and what is prevailing presently.

In fact competition amongst application software developers started way back in 1960s-70s when MRP (Material Requirement Planning) got developed to automate inventory and production business processes in some manufacturing plants. Eventually, several large software providers, including Oracle and JD Edwards (now taken over by Oracle), set out to make this software accessible to more businesses. The '80s marked a milestone in the history of application integration when the first manufacturing resource planning (MRP II) systems appeared to support manufacturing processes beyond inventory and raw materials procurement. MRP II systems allowed the various departments involved in manufacturing to coordinate, and they had more advanced production scheduling capabilities.

By 1990, Gartner, as stated earlier, introduced the term ERP to recognize that many business functions, like accounting, sales, engineering and human resources (HR) apart from manufacturing were now using this technology to increase the efficiency of their entire operations also facilitating all employees with a single source of accurate data. Cloud ERP, came into existence first by NetSuite in 1998 bringing an improvement over on-premises systems. With this, available with all the main IT service providers, organizations can access critical business data through the web from any device with an internet connection with no longer need to purchase and maintain hardware. It also reduces the need for IT staffers and implement easily. This cloud model has made ERP systems, once limited to enterprises, accessible to smaller companies that lack the capital to launch and support a resource-intensive on-premises solution. Small and midsize businesses across industries can enjoy the same benefits as their larger counterparts, including automated processes, improved data accuracy and greater efficiency.

Integrated application systems have come a long way and technological innovations will lead to continued advancements with integration of Artificial Intelligence (AI) into them.

There are more than 100 ERP systems available worldwide for Small Business, Mid-Sized, Fast-Growing, and Large Enterprise. A survey based on 129 products indicates that top five companies providing integrated applications are /Oracle, SAP, SAGE, SYSPRO, and Microsoft. (Source : <https://www.selecthub.com/erp-software/>).

For comparison, we will take up the integrated applications with Oracle, SAP, and Microsoft, most popular globally, with their diversified products .

8.7 BUSINESS SOFTWARE SOLUTIONS FROM ORACLE

The integrated applications under the umbrella of Oracle are Oracle's E-Business Suite, JD Edward's EnterpriseOne, Hyperion, Peoplesoft, and

Siebel. Oracle took over JD Edward in 2005. JD Edward, earlier to it, had taken over Peoplesoft.

Oracle's E-Business Suite (EBS) offers a complete set of applications, which are capable of automating entire business processes encompassing wide array of functions.

EBS Mobile Release 9 also updates to smart phone apps. EBS Cloud Manager helps enterprises to run EBS on Oracle Cloud Infrastructure (OCI).

Oracle EBS is an integrated set of business applications that deliver comprehensive procedural computerization and thus complete information. Information finally helps businesses in taking more informed decisions. Thousands of the world's leading organizations rely on Oracle E-Business Suite to run their key business operations. Oracle Cloud Infrastructure, helps you save money, gain productivity, and respond to changing business demands with automated, scalable cloud infrastructure. Table 8.1 gives you a global view of different modules of Oracle E-Business Suite. Apart from the EBS, Oracle also provides business technology platform for database and data management, application development and integration, Analytics, and intelligent technologies for doing businesses in state-of-the-art areas and applications.

Table 8.1 : Oracle solutions under its E-Business Suite

S. No.	Solutions/ Modules	Description
1	Financials	Oracle Financials is an integrated solution that gives you complete visibility into your finances and control over all transactions, while increasing the efficiency of your operations. It covers Receivables, Payables, Assets, Lease Contracts (Financials), Lease and Finance Management.
2	Human Capital Management	This solution readies your company for human resources to rapidly deploy the best for maximum employee productivity, satisfaction, and retention.
3	Business intelligence	Business intelligence solutions help you meet the information needs of your users by providing self-service, Web-based access to information. It provides hundreds of pre-built reports and analytic capabilities. Oracle addresses the full spectrum of business intelligence requirements, such as data quality, data analysis, and information access.
4	Learning Management	Learning Management module provides a complete infrastructure that lets organizations manage, deliver, and track training, in both online and classroom environments.
5	Maintenance Management	Maintenance Management module helps you plan and execute maintenance activity for any asset, as well as keep track of information on asset costs and activities throughout an enterprise.
6	Manufacturing	Manufacturing module is a set of collaborative applications that optimize production capacity beginning with raw materials through final products, regardless of the manufacturing methodology used with Discrete Manufacturing, Process Manufacturing, Outsourced Manufacturing, Cost Management, Quality etc.
7	Marketing	Oracle Marketing integrates with Sales, Finance and the rest of your organization, for better planning and execution. It drives profit by intelligently marketing to the most profitable customers.

8	Order Management	Order Management uses Oracle Workflow to support tailored, automated fulfillment processes, without the need for customization. It captures demand from multiple channels using a variety of communication methods with Order Management, Inventory Management, Advanced Pricing, Shopping Experience, Landed Cost Management, Channel Revenue Management, Incentive Compensation
9	Procurement	Procurement is an integrated set of applications that connects all your purchasing functions, including sourcing, purchase order management, receiving, and payables, to give you a central repository of purchasing information, along with fully automated, self-service capabilities for Shopping, Procurement etc.
10	Product Lifecycle Management	Oracle Product Lifecycle Management is collection of applications that gives you a centralized repository of product and project information that will help you manage product activities collaboratively.
11	Projects	Projects, a set of Internet-enabled applications, delivers centralized project management functionality and serves as a bridge between operations and corporate finance to help you effectively manage the full lifecycle of every project, Project Procurement, Contract Lifecycle Management etc.
12	Sales	Sales enables you to have comprehensive customer and prospect information so that your sales organizations can meet challenging revenue goals with fewer resources. .
13	Supply Chain Management	Supply Chain Management lets you gain global visibility, automate internal processes and readily collaborate with your suppliers, customers, and partners.
14	Asset Life Cycle and Service	It helps enterprise Asset Management, Asset Tracking, Service Contracts, Service (TeleService), Field Service, Depot Repair etc.

(Source: <https://www.oracle.com/a/ocom/docs/applications/ebusiness/announcing-eb-innovations-2020.pdf>)

8.8 BUSINESS SOFTWARE SOLUTIONS FROM SAP

SAP (System Analysis and Program Development) was founded on April 1, 1972 in Germany by five former IBM employees with an idea to create standard enterprise software that could integrate all business processes and enable data processing in real time. Their products gained prominence with the application suite “R/3” in 1992 with the client-server software smoothing the path to a globalized economy, turning SAP into a global player. The company branched out into three markets of the future: mobile technology, database technology, and cloud. SAP acquired some of its competitors, including Business Objects, Sybase, Ariba, SuccessFactors, Fieldglass, and Concur. SAP has more than 238 million cloud users currently, more than 100 solutions covering all business functions, and the largest cloud portfolio of any provider. SAP operates 70 data centers at 37 locations in 17 countries.

SAP HANA was launched in 2011 giving very fast access to customers. Data analyses that used to take days or even weeks were now completed in seconds. Four years later, SAP launched SAP S/4HANA, its latest generation of business software. In 2019, SAP acquired U.S. company Qualtrics, a

leader in experience management software, placing SAP at the leading edge of this growing segment. The company's integrated applications connect all parts of a business into an intelligent suite on a digital platform. SAP Business Technology Platform powers customers to become intelligent enterprises and supports cloud, on-premise, and hybrid customer landscapes.

Integrated Applications by SAP is the Business Suite which is a bundle of end-to-end enterprise software applications integrating data, processes and the business functions. SAP's solutions are available for many industries like aerospace & defense, engineering, construction, operations, financial service providers, insurance, telecommunications and banking etc. Business Suite is used primarily by medium to large enterprises. It helps to run businesses better, faster, and simpler enterprise-wide responding more quickly to changing customer needs and market conditions. It includes CRM, Financials, HR, SRM, SCM and many more. The table given below describes them. The foundation of Business Suite is SAP ERP, often referred to as ERP Central Component (ECC). The Table 8.2 gives you a global view of different modules of SAP ECC.

Table 8. 2: SAP solutions under its business suite SAP ECC

S. No.	Solutions/ Modules	Description
1	Financial Accounting (FI)	The FI component records a company's all financial transactions. It handles receivables from sales, payables for procurement, and cash management as well as bank payment and reconciliation processes.
2	Controlling (CO)	The CO component manages cost center accounting, profit center accounting and internal orders and also offers financial planning. It's product costing feature is used to compare simulated costs and actual costs primarily intended to help manufacturing organizations. Together FI and CO are often referred to as FICO, FI-CO or FI/CO.
3	Sales and Distribution (SD)	The SD component manages major processes of sales and distribution. This includes selling products or services in national and international markets through direct sales to customers or through distribution networks. It also handles customer's returns, along with billing and credit issuance.
4	Materials Management (MM)	The MM component manages procurement, from local vendors or international suppliers, and inventory for all goods issuance, goods receipts and transfers of a material from one plant or storage location to another. Counting and reconciling materials' physical inventory is also managed in MM.
5	Production Planning (PP)	The PP component helps businesses create demand and manufacturing capacity alignment to help plan product manufacturing, sales and distribution. It can be used for discrete, process or repetitive manufacturing.
6	Quality Management (QM)	The QM component extensively integrates with procurement, production, sales, and equipment maintenance processes. It can be used to manage complete internal or external audit's business processes. It can also assist in finding root causes of a product's failure to ensure ongoing quality improvements to a company's business processes.

7	Plant Maintenance (PM)	The PM component monitors machines and functional locations (such as a chiller room or a boiler room) to make sure that they are in proper working order providing alerts when issues are detected to prevent machine failures and production disruptions. Business processes such as preventive maintenance, corrective maintenance and refurbishment maintenance are all covered in this module.
8	Customer Service (CS)	The CS component handles business processes related to a company providing maintenance services to customers' equipment.
9	Project System (PS)	PS is meant to manage large, complex projects such as setting up a new manufacturing plant or monitoring a plant's maintenance turnaround. All project-specific procurement or production through PS ensures that this component is able to allocate a project's costs correctly while remaining within the defined project's budget.
10	Human Capital Management (HCM)	HR is used for payroll, time management activities such as attendance and leaves, career development, travel and workplace safety.
11	Customer Relationship Management (CRM)	It focuses on sales and marketing, SAP customer experience solutions, customer data, marketing, commerce, sales, and service
12	Supplier Relationship Management	It covers procure-to-pay and collaborative service procurement
13	Product Lifecycle Management	It helps to manage a company's product lifecycle needs
14	Supply Chain Management	It focuses on business planning, logistics and order fulfillment with supply chain planning, supply chain logistics, manufacturing, and enterprise asset management

(Source: <http://www.sap.com/>)

SAP also provides state-of-the-art business technology platform for database and data management, application development and integration, Analytics and intelligent technologies for digital transformation, machine learning, Blockchain, Internet of Things, Cloud Computing, Big Data. You will study about these terms for better understanding in further Units.

We will now have a look at the business solutions offered by Microsoft.

8.9 BUSINESS SOLUTIONS FROM MICROSOFT

Microsoft Business Solutions offers integrated business applications. It creates applications and services for retailers, manufacturers, wholesale distributors, and service companies like the other two players covered in earlier Sections. The Microsoft product which provides integrated business solutions is known today as Microsoft Dynamics 365. With Dynamics 365, customers have the only portfolio of intelligent business applications that empowers everyone to deliver operational excellence and create more engaging customer experiences. Dynamics 365 applications can be easily connected with the systems and tools which are already in use by customers

to add and extend capabilities for the existing capabilities in order to accelerate the business growth with optimized operations connecting with every customer in more innovative way with mixed reality and Internet of Things (IoT).

MS Dynamics 365 has about 12 standard application modules which can be integrated as one software solution. Based on the requirements customers can select one, some, or all application modules for the business growth. The modules are Finance and Operations, Retail, Human Resource, Sales, Customer Service, Project Service Automation, Field Service, Marketing, Supply Chain Management, Artificial Intelligence(AI) for Sales/Customer Service/Market Insight, Mixed Reality for Remote Assistance and Layout, Business Central as ERP for SMBs i.e. for Small and Middle level Businesses. Apart it has eCommerce module for online shopping, marketing, and eCommerce activities. Microsoft itself uses the Customer Service module to manage the process of updating Windows operating systems and other software resources, provided by them, to the latest versions. Further, MS Dynamics eShop Sales Portal is fully-integrated with a simple user interface built for in-house sales representatives. With eShop, sales team is sufficiently equipped with all what they need to maximize sales and provide great customer order management. Table 8.3 gives you a global view of different application software modules of Microsoft Dynamics 365.

Table 8.3: Microsoft's Integrated Applications

S. No.	Applications/ Modules	Description
1	Marketing	To create connected customer experiences, find and nurture new customers, stay on top of the market trends, strengthen relationships and earn loyalty, help with compliance requirements and protect customer data. Easily customize and connect with tools you already use. Personalize customer experiences with AI.
2	Sales	It can be used to find and build stronger relationship, improve productivity and performance, to get a single view of customers, explore e-books, videos, webinars etc.
3	Customer Service	To deliver positive customer experiences faster, optimize resources and help technicians to be more efficient, reduce operational costs, quickly address and resolve issues using AI-powered virtual agents, freeing agents to handle more complex matters. To deliver immediate access to the right content at the right time, collaborate, share tips, and leverage rich insights from experts and external users.
4.	Field Service	To ensure consistent and dependable operations by predicting, detecting, and resolving service issues before customers even know there's a problem. To deliver exceptional onsite customer experiences while optimizing resources and costs by dispatching technicians when and where they're needed most.
5.	Finance	To automate finance management processes to increase efficiency, decrease operational expenses and financial complexities, improve financial controls, and make timely decisions to drive agility and growth using comprehensive, real-time financial reporting, embedded analytics, and AI-driven insights, minimize costs and optimize spending across business geographies with process automation, budget control, and financial planning and analysis.

6	Supply Chain Management	To connect operations across small or medium-sized business, predict disruptions and respond fast by digitizing the supply chain, enhancing visibility, improving planning, and maximizing asset productivity.
7	Business Central	To ensure business continuity with a cloud solution that connects sales, service, finance, and operations teams to help them adapt faster and deliver results.
8	Project Operations	To connect project-centric business in one application, get the visibility, collaboration, and agility needed to drive success across project-centric business—from prospects to payments to profits.
9	Commerce	To unify physical and digital commerce, build brand loyalty through personal engagement, and exceed customer expectations.
10	Human Resources	To take the work out of managing the workforce, streamline and optimize processes, empower employees with skill development and set them up for success.
11	Manufacturing	To coordinate entire manufacturing process from product configuration, supply and capacity requirements planning, to scheduling and managing shop floor.
12	Retail Management	To run retail operations from point-of-sale to delivery, increase customer flow, speed up lines and tasks, control inventory, and automate purchasing.

(Source: <https://dynamics.microsoft.com/en-us/>)

So, you can see that there is a solution for virtually every thing that your business needs. For all type of business i.e. Small, Medium, or Large, integrated application solution is available worldwide with state-of-the-art technologies. It is up to the business houses to select the right solution for the respective requirements.

Activity B

Go to the websites of Oracle and SAP, list down five business advantages or competitive edge that their customers achieved after the implementation of their business solution.

.....

.....

.....

.....

8.10 SUMMARY

Integration of business solutions has emerged as a basic requirement of business today. No one wants to wait for the information coming from various departments and then compile for further use. Technology available today is well capable of providing right information at the right time. It is up to the user to decide what they want. Right from IT infrastructure to integrated application software, everything is available with no hurdle of devoting much time towards change over from the previous practices. Integration of business solution has emerged as an important tool here. They connect your customers, suppliers, employees any time and all the

time from any where. Thus it works for your whole business and not just one department. The unit has tried to explain the concepts of integration and has highlighted the solutions, which are available in the market today.

8.11 UNIT END EXERCISES

1. What are the technology related challenges that are invisible to your organization? Highlight the pitfalls? How integrated software applications can help in this?
2. What are integrated software applications? Talk about their advantages and business utility.
3. Write notes on the business software solutions provided by
 - a. SAP
 - b. Oracle
 - c. Microsoft
4. Compare the Business solutions provided by Microsoft and SAP. Clearly highlight the features which are common between the two.
5. Compare the Business solutions provided by Oracle and SAP. Clearly highlight the features which are common between the two.

8.12 REFERENCE & SUGGESTED FURTHER READINGS

<https://www.webopedia.com> , August 3, 2021

<https://en.wikipedia.org> , August 5, 2021

<https://www.g2.com/categories/erp-systems>, August 9, 2021

<https://www.netsuite.com/portal/resource/articles/erp/erp-history.shtml>, August 13, 2021

<https://terillium.com/comparing-types-of-erp-systems-for-your-organization/>, August 12, 2021

<https://www.selecthub.com/erp-software/> August 12, 2021

<https://www.oracle.com/a/ocom/docs/applications/ebusiness/announcing-ebs-innovations-2020.pdf> , August 15, 2021

[https://www. searchsap.techtarget.com](https://www.searchsap.techtarget.com), August 17, 2021

https://en.wikipedia.org/wiki/Microsoft_Dynamics_365 , August 18, 2021

<https://dynamics.microsoft.com/en-us/what-is-dynamics365/> , August 18, 2021

<https://www.dynamicseshop.com/integrated-mobile-solution-for-dynamics-365-ERP> , August 18, 2021

<http://www.sap.com/> for information about SAP Business Solutions

Block

3

SYSTEM ANALYSIS AND COMPUTER LANGUAGES

UNIT 9

Building Information Systems**145**

UNIT 10

System Analysis and Design**172**

UNIT 11

Computer Programming and Languages**187**

BLOCK 3: SYSTEM ANALYSIS & COMPUTER LANGUAGES

Till now you have read about the hardware, software, and networking technologies. You have also learned about the information systems in general and applications of information systems in functional areas in particular. This block tries to perceive you in the hat of a programmer or a system developer.

Unit 9: Building Information Systems make an effort to enlighten you with different types of Information system and their impact on doing business. You would be on familiar terms with planned organizational charge. You will understand the business values of Information System and appreciate the advantages and disadvantages of outsourcing Information System. You would also learn about how IS can be used to ensure quality.

Unit 10: System Analysis and Design strive to acquaint with traditional systems life cycles. You will identify different phases of system life cycle and understand system analysis. It will enable you to empathize with the different types of system design. This Unit also makes you aware of the latest concepts in systems design. You will also know about issues in implementation and maintenance of software.

Unit 11: Computer Programming & Languages highlights the concept, fundamentals and vocabulary of programming languages. You will grasp the important features of Visual Basic, Java, HTML & XML, and Excel, and would be able to write programs in higher-level language.

THE PEOPLE'S
UNIVERSITY

UNIT 9: BUILDING INFORMATION SYSTEMS

Structure

- 9.1 Introduction
- 9.2 Objectives
- 9.3 Computer Based information systems and its classification
 - 9.3.1 Office Automation Systems
 - 9.3.2 Communication Systems
 - 9.3.3 Transaction Processing systems
 - 9.3.4 Enterprise Information Systems
 - 9.3.5 Decision Support Systems
 - 9.3.6 Execution Systems
 - 9.3.7 Going beyond the information system categories
- 9.4 Redesigning the organisation with information systems
- 9.5 Business values of information systems
- 9.6 Outsourcing information system
 - 9.6.1 Advantages and Disadvantages of Outsourcing
 - 9.6.2 When to Use Outsourcing
- 9.7 Ensuring quality with information system
- 9.8 Summary
- 9.9 Unit End Exercises
- 9.10 References & Suggested Further Reading

9.1 INTRODUCTION

An Information System is a set of people, procedures, and resources that collects, transforms, and disseminates information in an organization. In the world over, today's end users rely on many types of Information Systems (IS). Automation has penetrated everywhere and in every field of life. In this unit, we will concentrate on *computer-based information systems* that use hardware, software, communication technology, and other forms of *Information Technology* (IT) to transform data resources into a variety of information products. Four kinds of organizational changes are enabled by Information systems. These are automation, rationalization, re-engineering and paradigm shift. When an organization does not use its internal resources to build and operate information system it takes help of other organizations to provide these services. This is called outsourcing. There are advantages and disadvantages of using outsourcing. Quality programs differ greatly from company to company. Some are merely generalized "sales" campaigns intended to sensitize employees to the need to strive for more quality in their daily work. At the opposite extreme, quality programs can result in

fundamental changes in the way a company does its business. Companies also follow different routes in achieving quality. Whatever route a company selects, the more it tries to achieve with its quality programs, the more information systems can contribute the success of those programs.

9.2 OBJECTIVES

After reading this unit, you should be able to

- Identify different types of Information Systems;
- Explain the impact of Information System on doing business;
- Describe about planned organizational change;
- Understand the business values of Information System;
- Enumerate the advantages and disadvantages of outsourcing Information System; and
- Discuss how IS can be used to ensure quality.

9.3 COMPUTER BASED INFORMATION SYSTEMS AND ITS CLASSIFICATION

A computer based information system uses the resources of people (end users and IS specialists), hardware (machines and media), and software (programs and procedures), to perform input, processing, output, storage, and control activities that convert data resources into information products as shown in Figure 9.1.

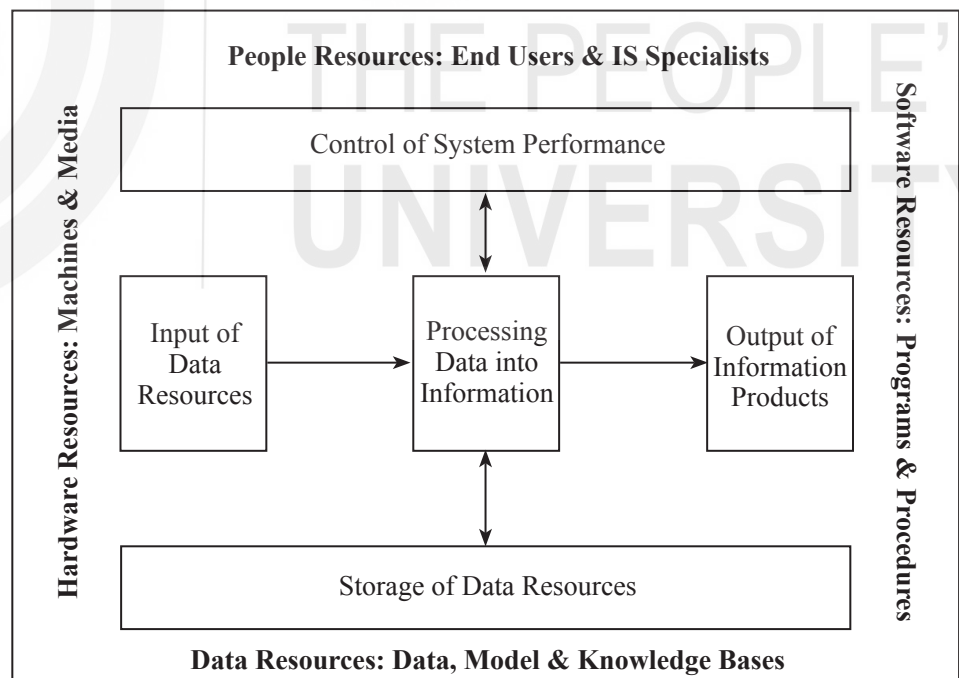


Figure 9.1: The components of an Information System

Source: James A, O'Brien, *Introduction to Information System*

Before going into the details of Computer based Information System, let us first discuss about a system. A system is a group of interrelated components working together toward a common goal by accepting inputs and producing outputs through an organized transformation process.

Such a system (sometimes called a dynamic system) has three basic interacting components or functions:

- Input involves capturing and assembling elements that enter the system to be processed. For example, raw materials, energy, data, and human effort must be secured and organized for processing.
- Processing involves transformation processes that convert input into output. Examples are a manufacturing process or mathematical calculations.
- Output involves transferring elements that have been produced by a transformation process to their ultimate destination. For example, finished products, human services, and management information must be transmitted to their human users.

The systems concept can be made even more useful by including two additional components: feedback and control. A system with feedback and control components is sometimes called a cybernetic system, that is, a self-monitoring, self-regulating system.

- Feedback is data about the performance of a system. For example, data about sales performance is feedback to a sales manager.
- Control involves monitoring and evaluating feedback to determine whether a system is moving toward the achievement of its goal. The control function then makes necessary adjustments to a system's input and processing components to ensure that it produces proper output. For example, a sales manager exercises control when he or she reassigns salespersons to new territories after evaluating feedback about their sales performance.

This information system model highlights the relationships among the components and activities of information systems. It provides a framework that emphasizes four major concepts that can be applied to all types of information systems:

- People, hardware, software, and data are the four basic resources of information systems.
- People resources include end users and IS specialists (like programmers, systems analysts, system and network experts etc), hardware resources consist of machines and media, software resources include both programs and procedures, and data resources can include data, model, and knowledge bases.
- Data resources are transformed by information processing activities into a variety of information products for end users.
- Information processing consists of input, processing, output, storage, and control activities.

Basic IS model shown in the given in Table 9.1. The table indicates that a computer based information system consists of four major resources: people, hardware, software, and data.

Table 9.1: Different resources of information systems

<u>People Resources</u> Specialists – systems analysts, programmers, system and network experts/operators. End users – anyone else who uses information systems. <u>Hardware Resources</u> Machines – computers, monitors, printers, optical scanners, switches, routers, hubs . Media – magnetic tape, optical disks, hard disk drives, USB drives, smart cards, and paper forms, cables etc. <u>Software Resources</u> Programs – operating system programs, worksheet programs, word processing programs, payroll programs, financial management program, integrated software programs (like ERP, CRM, SCM etc.) etc. Procedures – data entry procedures, error correction procedures, and paycheck distribution procedures etc. <u>Data Resources</u> Product descriptions, customer records, employee files, web sites, and databases etc. <u>Information Products</u> Management reports and business documents using text and graphics displays, audio & video responses, and paper forms etc.

There are many kinds of information systems in the real world. All of them use hardware, software, and people resources to transform data resources into information products.

It is important not to confuse information systems with the concept of computer systems. A computer system is a group of interconnected hardware components that may take the form of a microcomputer, minicomputer, or large mainframe computer system. However, whether it sits on a desk or is one of many computers in the internet or intranet network, a computer system still represents only the hardware resources component of a computer-based information system. As we have just seen, an information system also consists of people, software, and data resources.

Types of Information System

Although people often think of information systems as tools for decision-making, each type of information system supports both communication and decision-making in a number of ways.

Table 9.2: Typical ways each type of information system supports communication and decision making

System type	Typical user	Impact on communication	Impact on decision making
<i>Office automation system:</i> provides individuals effective ways to process personal and	Anyone who stores personal data, creates documents,	Provide tools for creating documents and presentations, such as work	Provides spreadsheets and other tools for analyzing information

organizational business data, to perform calculations, and to create documents	or performs calculation.	processors and presentation systems.	• Communication tools also help in implementing decisions.
<i>Communication system:</i> Helps people work together by sharing information in many different forms	• Anyone who communicates with others, including office workers, managers, and professions	• Telephones and teleconferencing for communication • Virtual conferencing and meetings • E-mail, v-mail, use of social media for communicating using messages and documents • Access to memos and other shared information • Scheduling meetings • Controlling flow of work	• Telephones and teleconferencing for decision making • Virtual conferencing and meetings • E-mail, v-mail, social media, other tools for obtaining information • Supports sharing information related to making joint decisions
<i>Transaction Processing System (TPS):</i> collects and stores information about transactions; controls some aspects of transactions	• People whose work involves performing transactions	• Creates a database that can be accessed directly, thereby making some person-to-person communication unnecessary	• Gives immediate feedback on decisions made while processing transactions • Provides information for planning and management decisions
<i>Management Information System (MIS) and Executive Information System (EIS):</i> converts TPS data into information for monitoring performance and managing an organization; provides executives information in a readily accessible interactive format	• Managers, executives, and people who receive feedback about their work	• Provides a basis of facts rather than opinions for explaining problems and their solutions • May incorporate e-mail and other communication methods with presentation of computerized data	• Provides summary information and measures of performance for monitoring results • May provide easy ways to analyze the types of information provided in less flexible form by older MIS
<i>Decision Support System (DSS):</i> Helps people make decisions by providing information, models, or analysis tools	• Analysis managers and other professionals	• Analysis using DSS helps provide a clear rationale for explaining a decision	• Provides tools for analyzing data and building models • Analysis using a DSS helps define and evaluate alternatives

<i>Execution system:</i> directly supports the organization's value added work (e.g., helps sales people sell, helps doctors practice medicine, or helps architects design buildings)	• People who do an organization's value added work, especially if that work involves special skills or knowledge	• May support communication or information sharing between people doing different parts of the task • May help explain the result of the task to customers	• May provide tools, information, or structured methods for making decisions • May store and provide expert knowledge to support decisions in specific areas
--	--	---	---

One of the reasons the various categories are mentioned frequently is that each is used in every functional area of business.

Table 9.3: Examples of each type of information system in three functional areas of business

System Type	Sales Examples	Manufacturing Examples	Finance Examples
Office automation systems	• Spreadsheet to analyze different possible prices • Word processor to create sales contract	• Spreadsheet to analyze a production schedule • Word processor to write a memo about how to fix a machine	• Spreadsheet to compare several loan arrangements. • Word processor to write a memo about new financial procedures
Communication systems	• E-mail and use of social media to contact customer • Video conference and use of social media to present new sales materials to sales force • Work flow system to make sure all sales steps are completed • System to coordinate all work on a complex sales contract	• E-mail and use of social media to discuss a problem with a new machine • Video-conference to coordinate, manufacturing and sales efforts • Work flow system to make sure engineering changes are approved	• V-mail and use of social media to communicate with bank about loan arrangements • Video conference to explain effect of financing on factory investments • Work flow system to make sure invoice approval precedes payment • System for exchanging the latest information related to lawsuit
Transaction Processing System (TPS)	• Point of sale system for sales transactions • Keeping track of customer contacts during a sales cycle	• Tracking movement of work in process in a factory • Tracking receipts of materials from suppliers	• Processing credit card payments • Payment of stock dividends and bond interest
Management information system (MIS) and Executive	• Weekly sales report by product and region	• Weekly production report by production and operation	Receivables report showing invoices and payments

Information System (EIS)	• Consolidation of sales projections by product and region • Flexible access to sales data by product and region	• Determination of planned purchases based on a production schedule • Flexible access to production data by product and operation	• Monthly financial plan consolidation • Flexible access to corporate financial plan by line item
Decision Support System (DSS)	• System helping insurance salespeople test alternatives • Marketing data and models to analyze sales	• System displaying current priorities for machine operator • Production data and models to analyze production results • Use of a GDSS to identify production problems	• System analyzing characteristics of customers who pay bills promptly • Stock database and models to help in selecting stocks to buy or sell
Execution system	• System to generate competitive bids • System to help salespeople suggest the best choice for the customer	• System to diagnose machine failures • System to transfer customer requirements to an automated machine cell	• System to support a loan approval process • System to find price inconsistencies between different equity markets.

Activity A

List down the major business activities of your organization. Describe what kind of information systems are already being used for these activities. Also suggest some applications of information systems in the activities that are still done manually.

.....

.....

.....

9.3.1 Office Automation Systems

An office automation system (OAS) facilitates everyday information processing tasks in offices and business organizations. These systems include a wide range of tools such as spreadsheets, word processors, and presentation packages. Although telephones, e-mail, v-mail, and social media can be included in this category, we will treat communication systems as a separate category.

OASs help people perform personal record keeping, writing, and calculation chores efficiently. Of all the system types, OASs and communication systems are the most familiar to students. Tools generally grouped within the OAS category include:

Spreadsheets are an efficient method for performing calculations that can be visualized in terms of the cells of a spreadsheet. Although spreadsheet programs seem second nature today, the first spreadsheet program was VisiCalc, which helped create the demand for the first personal computers in the late 1970s.

Text and image processing systems store, revise, and print documents containing text or image data. These systems started with simple word processor but have evolved to include desktop publishing systems for creating complex documents ranging from brochures to book chapters.

Presentation packages help managers develop presentations independently, instead of working with typists and technical artists. These products automatically convert outlines into printed pages containing appropriately spaced titles and subtitles. These pages can be copied directly onto transparencies or slides used in presentations.

Personal database systems and note-taking systems help people keep track of their own personal data (rather than the organization's shared data.) Typical applications include an appointment book and calendar, a to-do list, and a notepad.

When using these tools for personal productivity purposes, users can apply any approach they want because the work is unstructured. In these situations, some individuals use them extensively and enjoy major efficiency benefits, whereas others do not use them at all. The same tools can also be used for broader purposes; however, in which they are incorporated into larger systems that organizations use to structure and routinize tasks. For example, a corporate planning system may require each department manager to fill in and forward a pre-formatted spreadsheet whose uniformity will facilitate the corporation's planning process.

9.3.2 Communication Systems

Electronic communication systems help people work together by exchanging or sharing information in many different forms. Every day we come across a lot of communications through social media too such as WhatsApp, Facebook, Messenger, Twitter, LinkedIn etc. New communication capabilities have changed the way many businesses operate by making it possible to do many things at a distance that previously required being present in a specific location. This section groups these tools into four general categories. Teleconferencing systems make it possible to hold same-time, different-place meetings. Messaging systems make it possible to transmit specific messages to specific individuals or groups of individuals. Groupware systems start with messaging but go further by facilitating access to documents and controlling team-related workflow. Knowledge management systems facilitate the sharing of knowledge rather than just information.

Teleconferencing

The use of electronic transmission to permit same-time, different-place meetings is called teleconferencing. We can think of a traditional telephone call as a minimal teleconference, but the term is normally applied to other options including audio conferencing, audio graphic conferencing, and video conferencing.

The distinction between these approaches is related to the type of information that is shared. Audio conferencing is a single telephone call involving three or more people participating from at least two locations. If several people on the call are in the same office, they can all participate using a speakerphone, which includes a high-sensitivity microphone and a loudspeaker that can be heard by anyone in a room. Audio graphic conferencing is an extension of audio conferencing permitting dispersed participants to see pictures or graphical material at the same time. This is especially useful when the purpose of the meeting is to share information that is difficult to describe, organize, or visualize, such as a spreadsheet or model used to perform calculations under different assumptions. Video conferencing is an interactive meeting involving two or more groups of people who can see each other using television screens. The least expensive forms of video conferencing are tiny cameras and 4-inch screens add to telephones or separate video conferencing windows displayed on computer screens. In typical business video conferencing, remote participants appear on a television screen.

Video conferencing simulates a face-to-face meeting without requiring unnecessary travel, which absorbs time and energy, not to speak of the cost of airplane and hotel bills. However, the effectiveness of videoconferences decreases if the participants lack a prior social bond. For example, doing sales calls via videoconference might seem tempting but might not foster the personal relationship needed to succeed in many sales situations. On the other hand, many banks have begun to experiment with stripped-down branch offices that have no tellers but permit customers to open accounts by video conferencing with multilingual staffers in another state. This is happening in other fields too like getting driving licenses and similar business activities through online portals and Apps.

Messaging Systems

Different-time, different-place communication has been used for centuries in the form of books and letters. Messaging system make it possible to transmit specific messages to specific individuals or groups of individuals. They use technologies such as electronic mail, voice mail, and fax to make different-time, different-place communication more effective.

The use of computers to send and retrieve text messages or document addressed to individual people or locations is called electronic mail (e-mail). Each user is identified by is usually based on the person's name and also serves as the person's e-mail address. The sender uses a word processor to create a message and then addresses it to a distribution list. The distribution list might be an individual account name or a group of names, such as those for the sales department or everyone working on a particular project. The recipient can read the message immediately or can wait until it is convenient. The recipient e-mail message can save it, print it, erase it, or forward it to someone else. The recipient can also edit the message to extract parts to be saved, printed, or passed on.

E-mail is effective in many situations, such as permitting you to leave a message without going through an additional person who might garble it. With e-mail you can send a message to a person traveling away from

the office who can log onto a network using a laptop computer. If you are working on a memo or other document and want to get feedback from someone before you distribute it, you can use e-mail to send it to the person for a quick response. E-mail also allows you to send the same message to many individuals without having to contact them individually. For example, a product designer can respond with a good idea, the minimal effort of distributing the request is worth it.

There have been many innovative uses of e-mail to improve communication. People in large organizations have used it to bypass bureaucratic structures. For example, top managers sometimes bypass intermediate management levels by obtaining specific information directly from people throughout the organization. Some organizations have replaced the majority of their formal memos with informal e-mail that gets to the point directly. E-mail removes accents and permits non-fluent speakers to read a message several times that otherwise might be misunderstood in a phone conversation. It also helps them express their ideas more effectively than they might by using a telephone.

Groupware

Groupware helps teams work together by sharing information and by controlling internal workflows. Coined in the late 1980s the term groupware has attained wide recognition due to the increasing need for groups to work together more effectively at a distance as a result of downsizing and rapid organizational change. Products viewed as groupware are still new enough that their long-term direction is unclear even though the competitive need to work effectively in dispersed teams is greater than ever. This is being used in all fields like education, hospitals, agriculture, industries, offices etc.

Groupware goes beyond messaging by facilitative access to documents and controlling team-related workflow. Many groupware products are related to specific group related tasks such as project management, scheduling meetings (“calendar”), and retrieving data from shared databases. Lotus Notes, a prominent product in this category, is designed for sharing text and images and contains a data structure that is a cross between a table-oriented database and an outline. For example, a law firm in Seattle uses Lotus Notes to permit everyone working on a particular case to have access to the most current memos and other information about that case, even if they are traveling. Other companies use Lotus Notes to store and revise product information for salespeople selling industrial products, thereby replacing the massive three-ring binders they formerly lugged around. In similar way, many such products are available today with Google and other prominent IT service providers.

Yet other groupware functions are performed through computer conferencing, the exchange of text messages typed into computers from various locations to discuss a particular issue. When done through the Internet this is sometimes called a newsgroup. A computer conference permits people in dispersed locations to combine their ideas in useful ways even though they cannot speak to each other face-to-face. Any conference participant may be able to add new ideas, attach comments to existing messages, or direct comments to specific individuals or groups. Proponents

of computer conferencing recognize some disadvantages of working through computers but emphasize major advantages, such as preventing a single forceful individual from dominating a meeting. Also, because everything is done through a computer, a record of how ideas developed is automatically generated.

A different type of groupware product focuses primarily on the flow of work in office settings. These products provide tools for structuring the process by which information for a particular multi-step task is managed, transferred, and routed. A typical example is the approval of travel expenditure. In this case, one person must propose the expenditure and someone else must approve it. The workflow application is set up to make the approval process simple and complete. In effect, groupware is being used as a small transaction processing system for multistep transaction.

Intranets and Extranets

The widespread use of the World Wide Web has led many firms to apply the information sharing concepts of groupware on a much larger scale by creating an additional type of communication system, intranets and extranets. Intranets are private communication networks that use the type of interface popularized by the Web but are accessible only by authorized employees, contractors, and customers. They are typically used to communicate non-sensitive but broadly useful information such as recent corporate news, general product information, employee manuals, corporate policies, telephone directories, details of health insurance and other employee benefits, and calendars. In some cases employees can use intranets to access and change their personal choices regarding health insurance and other benefits. Once security issues are addressed adequately, intranets for accessing general-purpose corporate data may lead to widespread use of intranets as a front end to transaction processing systems and management information systems described in the following sections.

Extranets are private networks that operate similarly to intranets but are directed at customers rather than at employees. Extranets provide information customers' need, such as detailed product descriptions, frequently asked questions about different products, maintenance information, warranties, and how to contact customer service and sales offices. Much of this information was formerly difficult for customers to access because paper versions of it at the customer site became scattered and outdated. By using extranets, companies are making this type of information increasingly available at a single interactive site that is easy to navigate.

Knowledge Management

A final type of communication system is very different from systems that support real time communication or provide access to information. Today's leading businesses are increasingly aware that their employees' knowledge is one of their primary assets. In consulting companies and other organizations that rely heavily on unique competencies and methods, knowledge has more competitive significance than physical assets because the physical assets can be replaced or replenished more easily.

Knowledge management systems are communication systems designed to facilitate the sharing of knowledge rather than just information. As with

groupware, the idea of knowledge management is still emerging and is applied in many different ways in different firms. The computer applications underlying knowledge management systems are often built on technologies such as intranets, electronic mail, groupware, databases, and search engines. Functions supported by these technologies include codifying knowledge (such as best practices), organizing it in repositories for later access, finding knowledge (using search engines and other schemes), and providing organized ways to find people who have needed knowledge.

The human element is paramount in knowledge management. The companies with the best results to date stitch technologies together into a system that operates effectively and that is genuinely supported by the culture. For example, employee reviews in many consulting companies give significant weight to demonstrated contribution to internal knowledge management systems. This type of recognition is especially important if the firm's culture otherwise encourages hoarding of knowledge for personal advancement. In many cases, the most effective use of knowledge requires involvement of the person who is the expert. When a British Petroleum drilling ship in the North Sea encountered an equipment failure, it put the equipment in front of a video camera and used a satellite link to contact a drilling expert in Scotland. His rapid diagnosis of the problem prevented delays and a possible shutdown.

9.3.3 Transaction Processing systems

A transaction processing system (TPS) collects and stores data about transactions and sometimes controls decisions made as part of a transaction. A transaction is a business event that generates or modifies data stores in an information system. TPSs were the first computerized information systems. We encounter computerized TPSs frequently, including every time we write a cheque, use a credit card, or pay a bill sent by a company. A TPS used to record a sale and generate a receipt is primarily concerned with collecting and storing data. If the TPS validates a credit card or helps a clerk determine whether to accept a personal check, it also controls decisions made within the transaction.

TPSs are designed based on detailed specifications for how the transaction should be performed and how to control the collection of specific data in specific data formats and in accordance with rules, policies, and goals of the organization. Most contain enough structure to enforce rules and procedures for work done by clerks or customer service agents. Some TPSs bypass clerks and totally automate transactions; such as the way ATMs automate deposits and cash with draws. A well-designed TPS checks each transaction for easily detectable errors such as missing data, data values that are obviously too high or too low, data values that are inconsistent with other data in the database and data in the wrong format. It may check for required authorizations for the transaction. Certain TPSs such as airline reservation systems may automate decision-making functions such as finding the flight that best meets the customer's needs. Finally, when all the information for the transaction has been collected and validated, the TPS stores it in a standard format for later access by others.

As anyone knows who has tried to make a reservation when a computerized reservation system is down, organizations rely heavily on their TPSs.

Breakdowns disrupt operations and may even bring business to a complete halt. As a result, a well-designed TPS has backup and recovery procedures that minimize disruptions resulting from computer outages.

Batch versus Real Time Processing

The two types of transaction processing are batch and real time processing. With batch processing, information for individual transaction is gathered and stored but isn't processed immediately. Later, either on a schedule or when a sufficient number of transactions have accumulated, the transactions are processed to update the database. With real time processing, each transaction is processed immediately. The person providing the information is typically available to help with error correction and receives confirmation of transaction completion. Batch processing was the only feasible form of transaction processing when data were stored only on punched cards or tapes. Real time transaction processing requires immediate access to an online database.

Batch processing is currently used in some situation where the transaction data comes in on paper, such as in processing cheques and airline ticket stubs. A batch approach is also used for generating paychecks and other forms of paper output that will be distributed after a delay. Unfortunately time delays inherent in batch processing may cause significant disadvantages. The central database may never be completely current because of transactions received while the batch was being processed. Worse yet, batching the transactions creates built-in delays, with transactions not completed until the next day in some cases. Even systems with interactive user interfaces may include lengthy delays before transactions are completed. For example, weekend deposits into many ATMs are not posted to the depositor's account until Monday. Even though the ATM's user interface is interactive, the system in a larger sense doesn't perform real time processing.

Compared to batch processing, real time processing has more stringent requirements for computer response and computer uptime. As is obvious when a travel agent says "Sorry, the computer is down," the jobs and work methods of the people in the real time TPS are designed under the assumption that the system will be up and available.

9.3.4 Enterprise Information Systems

Many firms have tried to take transaction processing to a higher level by creating enterprise information systems that encompass the transaction processing done in the various functional silos. The idea of these efforts is to create unified databases that permit any authorized individual to obtain whatever information would be helpful in making decisions across the organization. In theory at least, having all this information in a unified database should improve decision-making. Enterprise information systems are quite controversial because the effort to create them is enormous. They involve much more than changing the format of databases. Often it is necessary to change business processes to suit the needs of the information system instead of vice versa. Nonetheless, many organizations have found that the integration resulting from this large investment seems to be worthwhile. The last part of this discussion explains why these information

systems are usually called Enterprise Resource Planning (ERP) systems even though planning is not their main focus.

Management and Executive Information Systems:

A management information system (MIS) provides information for an organization's managers. The idea of MIS predates the computer age. For example, as long ago as the middle 1500s, the Fogger family in Augsburg, Germany, had business interests throughout Europe and even into China and Peru. To keep in touch, they set up a worldwide news reporting service through which their agents wrote letters about critical political and economic events in their areas of responsibility. These letters were collected, interpreted, analyzed, and summarized in Augsburg and answered through instructions sent to the family's agents. This paper-based system encompassing planning, execution, and control helped the family move more rapidly in the mercantile world than their rivals. Instructions went out to the agents; the agents executed their work and the agents reported their results.

Computerized MIS generates information for monitoring performance, maintaining coordination, and providing background information about the organization's operation. Users include both managers and the employees who receive feedback about performance indicators such as productivity.

The concept of MIS emerged partly as a response to the shortcomings of the first computerized TPSs, which often improved transaction processing but provided little information for management. Computerized MISs typically extract and summarize data from TPSs to allow managers to monitor and direct the organization and to provide employees accurate feedback about easily measured aspects of their work. For example, a listing of every sale during a day or week would be extremely difficult to use in monitoring a hardware store's performance. However, the same data could be summarized in measures of performance, such as total sales for each type of item, for each salesperson, and for each hour of the day. The transaction data remains indispensable, and the MIS focuses it for management.

As part of an organization's formal control mechanisms, an MIS provides some structure for the comparatively unstructured task of management by identifying important measures of performance. The fact that everyone knows how performance is measured helps in making decisions and helps managers motivate workers.

From MIS to EIS

An executive information system (EIS) is a highly interactive system that provides managers and executives' flexible access to information for monitoring operating results and general business conditions. These systems are sometimes called executive support systems (ESS). EIS attempts to take over where the traditional MIS approach falls short. Although sometimes acceptable for monitoring the same indicators over time, the traditional MIS approach of providing prespecified reports on a scheduled basis is too inflexible for many questions executives really care about, such as understanding problems and new situations.

EISs provide executives with internal and competitive information through user-friendly interfaces that can be used by someone with little computer-related knowledge. EISs are designed to help executives find the information they need whenever they need it and in whatever form is most useful. Typically, users can choose among numerous tabular or graphical formats. They can also control the level of detail, the triggers for exception conditions, and other aspects of the information displayed. Most EISs focus on providing executives with the background information they need, as well as help in understanding the causes of exceptions and surprises. This leaves executives better prepared to discuss issues with their subordinates.

9.3.5 Decision Support Systems:

A decision support system (DSS) is an interactive information system that provides information, models and data manipulation tools to help make decisions in semi structured and unstructured situations where no one knows exactly how the decision should be made. The traditional DSS approach includes interactive problem solving direct use of models, and user-controllable methods for displaying and analyzing data and in formulating and evaluating alternative decisions. This approach grew out of dissatisfaction with the traditional limitations of TPS and MIS. TPS focused on record keeping and control of repetitive clerical processes. MIS provided reports for management but were often inflexible and unable to produce the information in a form in which managers could use it effectively. In contrast, DSSs were intended to support managers and professionals doing largely analytical work in less structured situation with unclear criteria for success. DSSs are typically designed to solve the structured parts of the problem and help isolate places where judgment and experience are required.

DSSs may report repetitive or non-repetitive decision-making. They support repetitive decision making by defining procedures and formats, but they still permit the users to decide how and when to use the system's capabilities. They support non-repetitive decision making by providing data, models and interface methods that can be used however the user wants. The broad spectrum of information systems with the DSS label range from general tools such as spreadsheets, data analysis, and graphics packages to highly customized simulation or optimization models focusing on a specific business situation.

OLAP and Data Mining

The use of online data analysis tools to explore large databases of transaction data is called online analytical processing (OLAP). The idea of OLAP grew out of difficulties analyzing the data in databases that were being updated continually by online transaction processing systems. When the analytical processes accessed large slices of the transaction database, they slowed down transaction processing critical to customer relationships. The solution was periodic downloads of data from the active transaction processing database into a separate database designed specifically to support analysis work. The separate database often resides on a different computer, which together with its specialized software is called a data warehouse. Downloading data to a data warehouse makes it possible to perform both transaction processing and analytical processing efficiently without mutual interference.

Data mining is the use of data analysis tools to try to find the patterns in large transaction databases such as the customer receipts generated in a large sample of grocery stores across the United States. Careful analysis of this data might reveal patterns that could be used for marketing promotions, such as a correlation between diaper sales and beer sales during the evening hours.

9.3.6 Execution Systems

The information system categories discussed so far are primarily oriented toward planning and control activities or toward general office and communication activities. What about systems designed to directly support people doing the value added work that customers care about, such as practicing medicine, designing buildings, or selling investments? Some people call these systems “functional area systems”. Because there is no generally accepted term for information systems that support value added work, we will call them execution systems. These systems have become much more important in the last decade as advances in computer speed, memory capacity, and portability made it increasingly possible to use computerized systems directly while doing value added work. Such systems help plastic surgeons design operation and show the likely results to their patients help lawyers find precedents relevant to lawsuits, and help maintenance engineers keep machines running.

Expert systems are a type of execution system that has received attention as an offshoot of artificial intelligence research. An expert system supports the intellectual work of professionals engaged in design, diagnosis, or evaluation of complex situations requiring expert knowledge in a well-defined area. Expert systems have been used to diagnose diseases, configure computers, analyze chemicals, interpret geological data, and support many other problem solving processes. This type of work requires expert knowledge of the process of performing particular tasks. Although these tasks may have some repetitive elements, many situations have unique characteristics that must be considered based on expert knowledge. Intellectual work even in narrowly defined areas is typically much less repetitive than transaction processing general office work.

9.3.7 Going beyond the information system categories

The field of IT moves so rapidly that terminology often fails to keep pace with innovation. The same problem occurs with information system classification. People identify a new type of system, such as DSS or EIS, and describe its characteristics are no longer as important or have become commonplace. Eventually many information systems contain characteristics from several system categories. Furthermore a system that fits in a category today may not fit once new features are added. Information systems that contain characteristics of several different categories can be called hybrid information systems.

In the mid-1990s the widespread adoption of a form of hybrid information system called an Enterprise Resource Planning (ERP) system became highly visible and somewhat controversial. As with many IT terms, the term ERP evolved out of an early form of DSS called Material Requirements Planning (MRP). These systems provide an integrated view necessary to coordinate purchasing and production scheduling activities.

ERP systems try to create an integrated database that spans the major activities in a company. Ideally, having production, sales, human resources, and finances data in the same database should make it easier to analyze the business and to coordinate decision-making. Software vendors such as SAP, Microsoft, Oracle and other such popular IT service providers currently sell ERP software. These facilities are also available today through web in Cloud technology. These vendors analyzed basic business processes such as purchasing and of the process variations they found. This design strategy makes their products enormously complicated. Just figuring out which of the many options to use often takes several hundred person-months of time. In many situations, departments must give up existing customized systems that address their unique problems in order to use the more general software and its integrated database.

9.4 REDESIGNING THE ORGANIZATION WITH INFORMATION SYSTEMS

One of the most important things to know about building a new information system is that this process is one kind of planned organizational change. Frequently, new systems mean new ways of doing business and working together. The nature of tasks, the speed with which they must be completed, the nature of supervision (its frequency and intensity), and who has what information about whom will all be decided in the process of building an information system. This is especially true in contemporary systems, which deeply affect many parts of the organization. System builders must understand how a system will affect the organization as a whole, focusing particularly on organizational conflict and changes in the locus of decision-making. Builders must also consider how the nature of work groups will change under the impact of the new system. Builders determine how much change is needed.

New information systems can be powerful instruments for organizational change; Figure 9.2 shows that there are four kinds of structural organizational change which are enabled by information technology: automation, rationalization, re-engineering, and paradigm shifts. Each carries different rewards and risks.

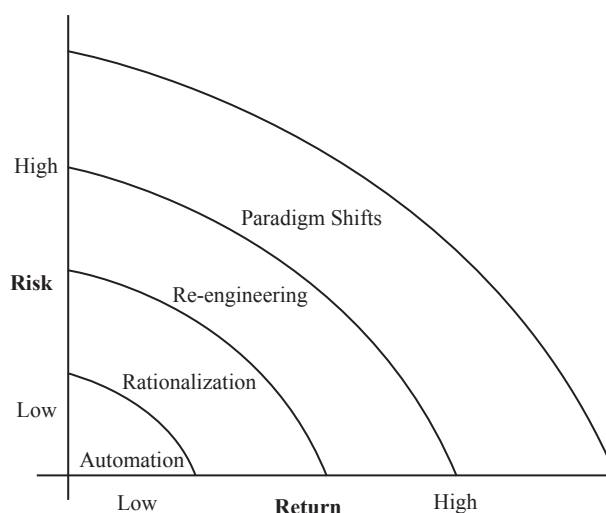


Figure 9.2: Organizational Change carries Risk & Rewards

Source: Kenneth C Laudon, Jane P Laudon, Prentice Hall India

The most common form of IT-enabled organizational change is automation. The first applications of information technology involved assisting employees perform their tasks more efficiently and effectively. Calculating paychecks and payroll registers, giving bank teller's instant access to customer deposit records, and developing a nationwide network of airline reservation terminals for airline reservation agents are all examples of early automation. Automation is akin to putting a larger motor in an existing automobile.

A deeper form of organizational change – one that follows quickly from early automation – is rationalization procedure. Automation frequently reveals new bottlenecks in production, and makes the existing arrangement of procedures and structures painfully cumbersome. Rationalization of procedures is the streamlining of standard operating procedures, eliminating obvious bottlenecks, so that automation can make operating procedures more efficient.

A more powerful type of organization change is business re-engineering, in which business processes are analyzed, simplified, and redesigned. Re-engineering involves radically rethinking the flow of work; the business procedures used to produce products and services with a mind of radically reduce the costs of business. A business process is a set of logically related tasks performed to achieve a defined business outcome. Some examples of business processes are developing a new product, ordering goods from a supplier, or processing and paying an insurance claim. Using information technology, organizations can rethink and streamline their business processes to improve speed, service and quality. Business re-engineering reorganizes workflows, combining steps to cut waste and eliminating repetitive, paper-intensive tasks (sometimes the new design eliminates jobs as well). It is much more ambitious than rationalization of procedures, requiring a new vision of how the process is to be organized. Rationalizing procedures and redesigning business processes are limited to specific parts of a business. New information systems can ultimately affect the design of the entire organization by actually transforming how the organization carries out its business or even the nature of the business itself.

This still more radical form of business change is called a paradigm shift. A paradigm shift involves rethinking the nature of the business and the nature of the organization itself. Banks, for instance, may decide not to automate, rationalize, or reengineering the jobs of tellers. Instead they may decide to eliminate branch banking altogether and seek less expensive source of funds, like international borrowing. Retail customers may be forced to use the Internet to conduct all their business, or a proprietary network. A paradigm shift is akin to rethinking not just the automobile, but transportation itself.

Of course nothing is free. Paradigm shifts and re-engineering often fail because extensive organization change is so difficult to orchestrate. Some experts believe that 70% of the time they fail. Why then do so many corporation entertain such radical change, because the rewards are equally high. In many instances firms seeking paradigm shifts and pursuing re-engineering strategies achieve stunning, order-of-magnitude increases in their returns on investment (or productivity).

9.5 BUSINESS VALUES OF INFORMATION SYSTEM

Another important fact about information systems is shown in Figure 9.3. No matter how they may be classified, information systems have following business values in an organization by supporting business operations, decision-making, and strategic management:

- Support of business operations.
- Support of managerial decision-making.
- Support of strategic competitive advantage.

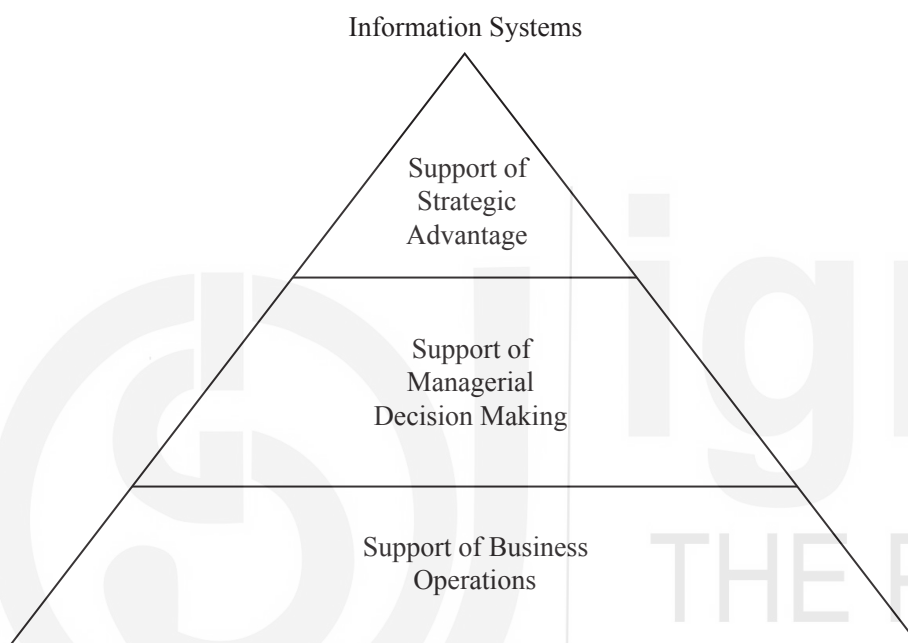


Figure 9.3: Business Values of Information System

Let's take a retail store as an example to illustrate this important first point. As a consumer, you have to deal regularly with information systems used to support business operations at the many retail stores where you shop. For example, most department stores use computer-based information systems to help them record customer purchases, keep track of inventory, pay employees, buy new merchandise, and evaluate sales trends. Store operations would grind to a halt without the support of such information systems.

Information systems also help store managers make better decisions and attempt to gain a strategic competitive advantage. For example, decisions on what lines of clothing or appliances need to be added or discontinued, or what kind of investments they require, are typically made after an analysis provided by computer-based information systems.

This not only supports the decision making of store managers but also helps them look for ways to gain an advantage over other retailers in the competition for customers. For example, store managers might make a decision to invest in a computerized touch-screen catalog ordering system as a strategic information system. This might lure customers away from

other stores, based on the ease of ordering merchandise provided by such a computer-based information system. Thus, strategic information helps provide strategic products and services that given an organization a comparative advantage over its competitors.

9.6 OUTSOURCING INFORMATION SYSTEMS

If a firm does not want to use its own internal resources to build and operate information systems, it can hire an external organization that specializes in providing these services to do the work. The process of turning over an organization's computer central operations, telecommunications networks, or applications development to external vendors of these services is called outsourcing.

Outsourcing information system is not a new phenomenon. Outsourcing options have existed since the dawn of data processing. As early as 1963, Petrot's Electronic Data Systems (EDS) handled data processing services for Frito-Lay and Blue Cross. Activities such as software programming, operation of large computers, time-sharing and purchase of packaged software have to some extent been outsourced since the 1960s.

Because information systems play such a large role in contemporary organizations, information technology now accounts for about half of most large firms' capital expenditure. In firms where the cost of information systems function has risen rapidly, managers are seeking ways to control those costs and are treating information technology as a capital investment instead of an operating cost of the firm. One option for controlling these costs is to outsource.

9.6.1 Advantages and disadvantages of outsourcing

Outsourcing is becoming popular because some organization perceive it as being more cost effective than it would be to maintain their own computer center and information systems staff. The provider of outsourcing services can benefit from economies of scale (the same knowledge, skills, and capacity can be shared with many different customers) and is likely to charge competitive prices for information systems services. Outsourcing allows a company with fluctuating needs for computer processing to pay for only what it uses rather than to build its own computer center to stand underutilized when there is no peak load. Some firms outsource because their internal information systems staff cannot keep pace with technological change. But not all organizations benefit from outsourcing, and the disadvantages of outsourcing can create serious problems for organizations if they are not well understood and managed.

Advantages of Outsourcing

The most popular explanations for outsourcing are the following:

Economy: Outsourcing vendors are specialists in the information systems services and technologies they provide. Through specialization and economies of scale, they can deliver the same service and value for less money than the cost of an internal organization.

Service quality: Because outsourcing vendors will lose their clients if the service is unsatisfactory, companies often have more leverage over external

vendors than over their own employees. The firm that out-sources may be able to obtain a higher level of service from vendors for the same or lower costs.

Predictability: An outsourcing contract with a fixed price for a specified level of service reduces uncertainty of costs.

Flexibility: Business growth can be accommodated without making major changes in the organization's information systems infrastructure. As information technology permeates the entire value chain of a business, outsourcing may provide superior control of the business because its costs and capabilities can be adjusted to meet changing needs.

Making fixed costs variable: Some outsourcing agreements, such as running payroll, are based on the price per unit of work done (such as the cost to process each cheque). Many out-sources will take into account variations in transaction processing volumes likely to occur during the year or over the course of the outsourcing agreement. Clients only need to pay for the amount of services they consume, as opposed to paying a fixed cost to maintain internal systems that are not fully utilized.

Freeing up human resources for other projects & financial capital: Scarce and costly talent within an organization can refocus on activities with higher value and payback than they would find in running a technology factory. Some agreements with outsource include the sale for cash of the outsourced firm's technology capital assets to the vendor.

Disadvantages of Outsourcing

Not all organizations obtain these benefits from outsourcing. There are dangers in placing the information systems functions outside the organization. Outsourcing can create serious problems such as loss of control, vulnerability of strategic information, and dependence on the fortunes of an external firm.

Loss of Control: When a firm farms out the responsibility for developing and operating its information systems to another organization, it can lose control over its information systems function. Outsourcing places the vendor in an advantageous position where the client has to accept whatever the vendor does and whatever fees the vendor charges. If a vendor becomes the firm's only alternative for running and developing its information systems, the client must accept whatever technologies the vendor provides. This dependency could eventually result in higher costs or loss of control over technological direction.

Vulnerability of Strategic Information: Trade secrets or proprietary information may leak out to competitors because a firm's information systems are being run or developed by outsiders. This could be especially harmful if a firm allows an outsourcer to develop or to operate applications that give it some type of competitive advantage.

Dependency: The firm becomes dependent on the viability of the vendor. A vendor with financial problems or deteriorating services may create severe problems for its clients.

9.6.2 When to use outsourcing?

Since outsourcing has both benefits and liabilities and is not meant for all organizations or all situations, managers should assess the role of information systems in their organization before making an outsourcing decision. There are a number of circumstances under which outsourcing makes a great deal of sense:

- When there is limited opportunity for the firm to distinguish itself competitively through a particular information systems application or series of applications. For instance, both the development and operation of payroll systems are frequently outsourced to free the information systems staff to concentrate on activities with a higher potential payoff, such as customer service or manufacturing systems. Applications such as payroll or cafeteria accounting, for which the firm obtains little competitive advantage from excellence, are strong candidates for outsourcing. If carefully developed, applications such as airline reservations or plant scheduling could provide a firm with a distinct advantage over competitors. The firm could lose profits, customers, or market share if such systems have problems. Applications where the rewards for excellence are high and where the penalties for failure are high should probably be developed and operated internally.

Companies may also continue to develop applications internally while outsourcing their computer center operations when they do not need to distinguish themselves competitively by performing their computer processing onsite.

When the predictability of uninterrupted information systems service is not very important. For instance, airline reservations or catalog shopping systems are too “critical” to be trusted outside. If these systems failed to operate for a few days or even a few hours, they could close down the business. On the other hand, a system to process employee insurance claims could be more easily outsourced because uninterrupted processing of claims is not critical to the survival of the firm.

When outsourcing does not strip the company of the technical know-how required for future information systems innovation. If a firm outsource some of its system but maintains its own internal information systems staff, it should ensure that its staff remains technically up to date and has the expertise to develop future applications.

When the firm’s existing information systems capabilities are limited, ineffective, or technically inferior. Some organizations use outsourcers as an easy way to revamp their information systems technology. For instance, they might use an outsourcer to help them make the transition from traditional mainframe-based computing to a new information architecture – distributed computing environment.

Despite the conventional wisdom on when to outsource, companies sometimes do outsource strategic functions. In any case, if systems development and the information systems function are well managed and productive, there may not be much immediate benefit that can be provided by an external vendor.

To obtain value from outsourcing, organizations need to make sure the process is properly managed. With sound business analysis and an understanding of outsourcing's strengths and limitations, managers can identify the most appropriate applications to outsource and develop a workable outsourcing plan.

Segmenting the firm's range of information systems activities into pieces that potentially can be outsourced makes the problem more manageable and also helps companies match an outsourcer with the appropriate job. Noncritical applications are usually the most appropriate candidates for outsourcing. Firms should identify mission-critical applications and mission-critical human resources required to develop and manage these applications. This would allow the firm to retain its most highly skilled people and focus all of its efforts on the most mission-critical applications development. Setting technology strategy is one area that companies should not abdicate to outsourcers. This strategic task is best kept in-house. Ideally, the firm should have a working relationship of trust with an outsourcing vendor. The vendor should understand the client's business and work with client as a partner, adapting agreements to meet the client's changing needs.

Firms should clearly understand the advantages provided by the vendor and what they will have to give up to obtain these advantages. For lower operating costs, can the client live with a five-second-response time during peak hours or next-day repair of microcomputers in remote offices? Organizations should not abdicate management responsibility by outsourcing. They need to manage the outsourcer as they would manage their own internal information systems department by setting priorities, ensuring that the right people are brought in, and guaranteeing that information systems are running smoothly. They should establish criteria for evaluating the outsourcing vendor that include performance expectations and measurement methods for response time, transaction volumes, security, disaster recovery, backup in the event of a catastrophe, processing requirements of new applications and distributed processing on microcomputers, workstations, and LANs. Firms should design outsourcing contracts carefully so that the outsourcing services can be adjusted if the nature of the business changes.

Activity B

Take the case of a petrol pump, a bookstore, a software development company, and an electronic goods manufacturer. Is outsourcing information systems required for them? Describe what can be outsourced and the advantages.

.....
.....
.....

9.7 ENSURING QUALITY WITH INFORMATION SYSTEM

The emergence of a global economy has stimulated worldwide interest in achieving quality. Companies can no longer be satisfied with producing

goods and services that compete only with goods produced within their own country – consumers can now select from a broad range of products and services produced anywhere in the world. Before examining how information systems can contribute to quality throughout the organization, we must first define the term quality.

Traditional definitions for quality have focused upon the conformance to specifications (or the absence of variation from those specifications). With this definition, a producer can easily measure the quality of its products. Achieving quality under this definition requires three steps from the manufacturer: First, establish product specifications. Second, measure products as they are produced to determine whether or not they achieve the standards established in the specifications. Third, alter the manufacturing process whenever necessary to bring the products up to standard.

However, achieving quality is not quite that simple and direct. The definition of quality has been changing and broadening in recent years. Defining quality as conformance to specifications view it from a producer's perspective only. Customers have a different perspective, being more concerned with value for their Rupees. They normally apply three criteria. First, customers are concerned with the quality of the physical product. They want to know if the product is durable, how safe it is, its reliability, its ease of use and installation, its stylishness, and how well the producer supports the product. Second, customers are concerned with the quality of service, by which they mean the accuracy and truthfulness of the advertising, the timeliness and accuracy of the billing process, responsiveness to warranties (implied as well as specified), and ongoing product support. Finally, customer concepts of quality include the psychological aspects: how well do the sales and support staff know their products, the courtesy and sensitivity of the staff, and even their neatness, the reputation of the product. For companies to compete globally, they need to include a customer perspective in any definition of quality.

Today more and more businesses are turning to an idea known as total quality management. Total quality management (TQM) is a concept that makes quality the responsibility of all people within an organization. TQM holds that the achievement of quality control is an end in itself. Everyone is expected to contribute to the overall improvement of quality – the engineer who avoids design errors, the production worker who spots defects, the sales representative who presents the product properly to potential customers, and even the secretary who avoids typing mistakes. Total quality management encompasses all of the functions within an organization. TQM is based on quality management concepts developed by American quality experts. Japanese management adopted the goal of zero defects, focusing on improving their products or services prior to shipment rather than correcting them after they have been delivered. Japanese companies often give the responsibility for quality consistency to the workers who actually make the product or service, as opposed to a quality control department. Studies have repeatedly shown that the earlier in the business cycle a problem is eliminated, the less it costs for the company to eliminate it. Thus the Japanese quality approach not only brought a shift in focus to the workers and an increased respect for product and service quality but also lowered costs.

Information systems can fill a special role in corporate quality programs for a number of reasons. First, IS is deeply involved with the daily work of other departments throughout the organizations. IS analysis usually have taken a leading role in designing, developing, and supporting such varied departmental systems as corporate payrolls, patent research systems, chemical process control systems, logistics systems, and sales support systems. IS professionals also maintain their knowledge of these departments through their participation in departmental information planning. In addition, IS personnel are usually key to the sharing of data between departments because they have unique knowledge of the relationships between various departments. Often, only IS personnel know where certain data originate, how other departments use and store them, and which other functions would benefit from having access to them. With this broad understanding of the functional integration of the corporation, IS personnel can be valuable members of any quality project team.

The IS staff in effective information systems departments have three skills that are critical to the success of a quality program. First, they are specialists in analyzing and redesigning business processes. Second, many IS technicians are experienced in quantifying and measuring procedures and critical activities in any process. Typically, IS departments have long been involved with measurements of their own manager training has long been a staple of better IS departments; such training includes the use of project management, software. These skills can contribute a great deal to any serious quality program, which will normally be organized as a project and will usually be heavily task-oriented.

The information systems staff is the source of ideas on the application of technology to quality issues; often they are also the people who can make that technology available to the quality project. For example, with the help of IS departments, statistical analysis software is becoming more widely used in the drive for quality.

Benchmark: Many companies have been effective in achieving quality by setting strict standards for products, services and other activities, and then measuring performance against those standards. Companies may use external industry standards, standards set by other companies, internally developed high standards, or some combination of the three.

IS contributes to these efforts in many ways: IS staff participates in re-engineering projects and helps to design and build the systems that make the quality processes possible. Any study of quality programs shows that information is a top concern to those involved, and IS is often central to the collection of that information. To improve production or sales, for example, management needs data to determine both what is being done right and what is being done wrong. IS is usually the key to making that information available in a timely fashion and in a format useful to those who need it for quality purposes. For instance, manufacturing data have traditionally been supplied to management in summary form at the end of the manufacturing process. In effect it is historical data that at best can be used to reduce future problems. To provide better information for benchmarking, information

systems specialists can work with business specialists either to design new systems or to analyze quality-related data found in existing systems.

Use customer demands as a guide to improving products and services: Improving customer service, making customer service the number one priority, will improve the quality of the product itself.

Reduce cycle time: Experience indicates that the single best way to address quality problems is to reduce the amount of time from the beginning of a process to its end (cycle time). Reducing cycle time usually results in fewer steps, an improvement right there. But reducing cycle time has other advantages. With less time passing between beginning and end, workers will be better aware of what came just before, and so are less likely to make mistakes.

Improve the quality and precision of the design: Quality and precision in design will eliminate many production problems. Computer-aided design (CAD) software has made dramatic quality improvements possible in a wide range of businesses from aircraft manufacturing to production of razor blades.

Increase the precision of production: For many products, one key way to achieve quality is to tighten production tolerance. CAD software has also made this possible. Most CAD software packages include a facility to translate design specifications into specifications both for production tooling and for the production process itself. In this way, products with more precise designs can also be produced more efficiently.

Include line workers in any quality process: Experience has shown that involvement of the people who perform the function is critical to achieving quality in that function. Although the information systems are could potentially make many more contributions like these, its involvement in corporate quality programs has provoked a great deal of controversy. IS has been criticized for a reluctance to become involved in organization-wide quality programs. Often IS focuses exclusively upon technological capabilities while not reaching out to aid the rest of the company in the ways described above. For example, many IS departments are criticized for failure to use customer demands as a guide to improving their products and services. On the other hand, non-IS departments often fail to consider contributions the IS staff might make to their quality project and so do not reach out to involve them. It is not uncommon for IS to be viewed only as technical support with little to contribute to the planning or content of the quality program.

9.8 SUMMARY

In this unit we discuss about computer based information system and their different types. Information systems have been used by organization as an effective way for decision-making and in supporting communication. The field of IT moves so rapidly that terminology often fails to keep pace with innovation. The same problem occurs with information system classification. People identify a new type of system, such as DSS or EIS, and describe its characteristics are no longer as important or have become commonplace. Eventually many information systems contain characteristics

from several system categories. Furthermore a system that fits in a category today may not fit once new features are added. Information systems that contain characteristics of several different categories can be called hybrid information systems.

Organizations in the past few years have shown a tendency to focus on their core business. Functions that are not considered to be part of the core business are outsourced to external suppliers. Organizations face an increased need, flexibility as a consequence of a faster changing competitive environment and rapid developments in information technology. Management has been questioning the idea that quality costs more. Today many senior executives have come to the conclusion that the lack of quality is actually a significant expense. While we all understand that product returns and repairs result in added costs for repair (labor, parts replacement, and additional shipping), only recently has management focused on the many previously hidden costs that arise from producing products that are not high quality.

9.9 UNIT END EXERCISES

- Q1. What types of resources are used in an information system?
- Q2. What are the different types of computer based information system used in different functional areas business by organizations?
- Q3. Who are the typical users of information system?
- Q4. What is the most common form of IT enabled organizational change?
- Q5. Which form of organizational change requires business processes to be analyzed simplified and redesign?
- Q6. When outsourcing should be used by the organizations?
- Q7. How Information systems can contribute for total quality management?

9.10 REFERENCES & SUGGESTED FURTHER READING

- 1. Alter Steven , *Information Systems (A management perspective)*, Pearson Education.
- 2. Burch, John , *Systems Analysis, Design, and Implementation*, Boston: Boyd & Fraser.
- 3. Laudon Kenneth, C., Laudon Jane, P. , *Management Information Systems (Organization and Technology)*, Prentice Hall of India, New Delhi.
- 4. Loff De , *Information Systems Outsourcing Decision Making (A Managerial Approach)*, IDEA Group.
- 5. O' Brien, James, A. *Introduction to Information Systems*, Irwin.
- 6. Post Gerald, V., and Anderson David, L.(2003), *Management Information Systems*, Tata McGraw-Hill, New Delhi.

UNIT 10: SYSTEM ANALYSIS AND DESIGN

Structure

10.1 Introduction

10.2 Objectives

10.3 Application Software Development and Customisation

10.4 Systems Development Life Cycle (SDLC)

10.5 Phases of SDLC

10.5.1 Project Identification and Selection

10.5.2 Project Initiation and Planning

10.5.3 Analysis

10.5.4 Logical Design

10.5.5 Physical Design

10.5.6 Implementation

10.5.7 Operation and Maintenance

10.6 Summary

10.7 Unit End Exercises

10.8 References & Suggested Further Reading

10.1 INTRODUCTION

Software is required to automate a process. The process to be automated could be any part of an activity or many activities at the organizational level in terms of an Information system. Success of any system depends on the approach of building it. If the development approach is right, the system will work successfully. From the inception of an idea for a software system, until it is implemented and delivered, and even after that, the system undergoes gradual development and evolution. The software is said to have a life cycle composed of several phases. Each of these phases results in the development of either a part of the system or something associated with the system, such as a test plan or a user manual. In the traditional life cycle model, each phase has well-defined starting and ending points, with clearly identifiable deliverables to the next phase.

10.2 OBJECTIVES

After reading this unit, you should be able to

- Describe systems development life cycles;
- Identify different phases of systems life cycle;
- Explain the process of system analysis;
- Describe the conceptual basis of system design; and
- Know about Implementation and maintenance of software.

10.3 APPLICATION SOFTWARE DEVELOPMENT AND CUSTOMISATION

Information system is an arrangement of people, data, processes, information presentation and information technology that interacts to support and improve day-to-day operations of a business as well as support the problem solving and decision making needs of management and users.

Information system is categorised as application software. It can be standalone module i.e. subsystem or an integrated system to automate processes in the organisation. Software is developed using computer language(s). This process is also known as coding. Development of software is either done by internal team or it can be developed by any IT service provider as decided by the organisation.

Before going for the development of the software, all users of the desired software in the organisation give the requirements. Based on these requirements only the software is developed. Nothing more nothing less is the mantra which is followed for automating the business processes. Similar like constructing a building the software development undergoes several well defined steps like requirements gathering, doing analysis for the cost involved and readiness to adopt the new system, doing analysis for inter and intra process relationships, preparing designs, coding and testing of the developed software, installation of the software and finally putting the software into practice by imparting all necessary and sufficient training to users for making it sustainable. All these steps are linked to one another to build the application software. They have been elaborated in further sections.

You have seen the concept of integrated applications with examples like ERP, SCM, CRM etc. These software are not developed again and again by organisations for business automation. Rather they are developed by IT companies in such a way that the produced system becomes generic. They can be further changed to suit the requirements of organisations globally. Software tools are available to make such application software. The process of making changes in already developed application software is called customisation. It is standard practice available these days. Here also requirements from users are required to make changes accordingly. Coding is also required to fill gaps under the customisation process. Similarly other steps of software development are also required in adopting and adapting the system – be it CRM, SCM, ERP or any other third party developed software.

10.4 SYSTEMS DEVELOPMENT LIFE CYCLE

The goal of the information system life cycle is to keep the project under control and assure that the system produced satisfies the requirements. The systems life cycle is the oldest method for building information systems and is still used today for complex, medium or large systems projects. This methodology assumes that an information system has a life cycle similar to that of any living organism, with a beginning, middle, and an end. This is known as the systems development life cycle (SDLC). The

traditional system life cycle divides the project mainly into four steps, each of which has distinct deliverables, such as computer programs, standard operating procedures (SOPs), or design documents. The deliverables are related because each subsequent step builds on the conclusions of previous steps. The four steps in traditional SDLC are – Initiation, Development, Implementation, and Operations & Maintenance, as shown in Figure 10.1.

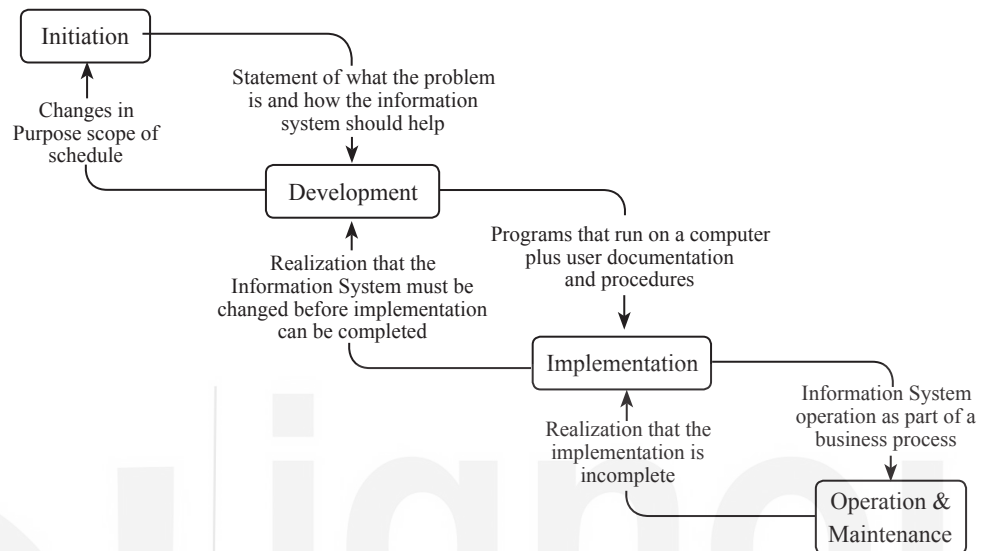


Figure 10.1: Traditional SDLC Phases

Some deliverables are oriented toward the technical staff, whereas others are directed toward or produced by users and managers. The latter ensure that users and their management are included in the system development process. These steps are covered in detail in next Section 10.4.

The traditional system life cycle is a tightly controlled approach designed to reduce the likelihood of mistakes or omissions. Despite its compelling logic, it has both advantages and disadvantages. Adherence to fixed deliverables and signoffs improves control but guarantees a lengthy process. Having specific deliverables due at specific times makes the schedule of deliverables. When merely going through the motions of producing deliverables on schedule, participants may be tempted sometimes to turn in work that is incomplete and to approve documents they do not truly understand. However, the traditional SDLC is the standard against which other approaches are compared. Project managers who want to bypass some of its steps still need a way to deal with the issues they raise.

SDLC is a standard methodology for the development of Information System. Traditional SDLC was strictly sequential. The software developers had to first complete the previous phase then start the next phase. But now concept of Repository is introduced in SDLC and it is known as Fast SDLC methodology where work is done across shared repository. It means that all the inputs and outputs of phases must be stored in the repository. At any time developers can backtrack to previous phase and they can also work on two phases simultaneously. Repository is thus defined as data store of accumulated system knowledge i.e. system models, detailed specifications,

and any other documentation that has been accumulated during the system's development. This knowledge is reusable and critical to the production system's ongoing support. The repository is implemented with various automated tools and it is often centralized as an enterprise business and IT resource. In SDLC, it is also possible to complete some activities in one phase in parallel with some other activities of another phase. Sometimes, life cycle is iterative; that is, phases are repeated as required until a satisfactory and acceptable system is found. Some people consider life cycle to be spiral, in which we constantly cycle through the phases at different levels of detail. The life cycle can also be thought of a circular process in which the end of the useful life of one system leads to the beginning of another project that will develop a new version or replace an existing system altogether.

Every custom software producer will have its own specific detailed life cycle or system development methodology. Even if a particular methodology does not look like cycle, you will discover that many of SDLC steps are performed; SDLC techniques and tools are used. In order to make this unit generic, we follow a rather general life cycle model, as described in Figure 10.2. This model resembles a staircase with arrows connecting each step to the step before and to the step after it. This representation of the system development life cycle is sometimes referred to as the "waterfall model". We use this SDLC as one example of methodology but more as a way to arrange the steps of systems analysis and design. Each phase has specific outputs and deliverables that feed important information to other phases. At the end of each phase, system development project reaches a milestone and, as deliverables are produced, parties outside the project team often review them.

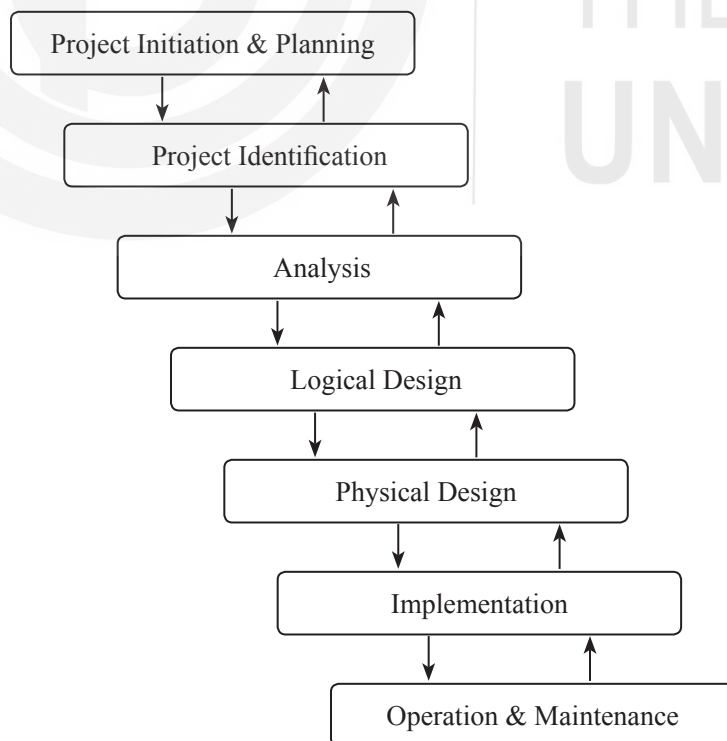


Figure: 10.2: Phases of System Development Life Cycle

Activity A

Take a business problem concerning your top manager and the IS department. Propose a system, which can be put on in such cases. Analyze the scope and feasibility of the proposed system and develop a project plan.

.....

.....

.....

.....

SDLC consists of mainly seven phases. These phases are also called as SDLC steps or stages. These phases/steps/stages are:

1. Project Identification and Selection
2. Project Initiation and Planning
3. Analysis
4. Logical Design
5. Physical Design
6. Implementation
7. Operation and Maintenance

10.5.1 Project Identification and Selection

The first phase in the SDLC is called project identification and selection. In this phase, the user identifies the need for a new or improved system. This phase stage tries to answer the questions, “Why do we need a new system ?” and “What do we want to accomplish?” This stage determines whether the organization has a problem and whether that problem can be solved by building a new information system or by modifying an existing one. If a system project is called for, this stage identifies its general objectives, specifies the scope of the project, and develops a project plan that can be shown to management. In large organizations, this identification may be part of a systems planning process. Information requirements of the organization as a whole are examined, and projects to meet these requirements are proactively identified. The organization’s information system requirements may result from requests to deal with problem in current system’s procedures, from the desire to perform additional tasks, or from the realization that information technology could be used to capitalize on an existing opportunity. These needs can then be prioritised and translated into a plan for the Information System department including a schedule for developing new major systems. In smaller organizations, determination of which systems to develop may be affected by user request submitted as the need for new or enhanced systems arises as well as from a formal information planning process. In either case, during project identification and selection, an organization determines whether or not resources should be devoted to the development or enhancement of each information system under consideration. The outcome of the project identification and selection process is a determination of which systems development projects should be undertaken by the organization at least in terms of an initial study.

10.5.2 Project Initiation and Planning

The second phase is project initiation and planning. The problems that are identified should be investigated and a decision to implement the information system or not for the organization should be taken. A critical step at this point is determining the scope of the proposed system. The project leader and initial team of system analysts also produce a specific plan for the proposed project, which the team will follow using the remaining SDLC steps. Now, this baseline project plan customizes the standardized SDLC and specifies the time and resources needed for its execution. The formal definition of a project is based on the likelihood that the organization's information system department is able to develop a system that will solve the problem or use the opportunity and determine whether the costs of developing the system outweigh the benefits it could provide, based on cost benefit analysis. In case of large projects the IS department may propose for getting the system developed and implemented by a third party. The final presentation with the subsequent project phases is usually made by the project leader and other team members to the organisation's management and/or to a special management committee with experts, some of them may be from other similar organizations. With all such presentations finally the decision is taken about the projects the organization will undertake.

10.5.3 Analysis

Analysis is the next phase. During this phase, the analysis has several sub-phases. The first is requirements determination. In this sub-phase, system analysts work with users to determine the requirements and expectations of users from the proposed system. The systems study stage analyzes the problems of existing systems (manual or automated) in detail, identifies objectives to be attained by a solution to these problems, and describes alternative solutions. The systems study stage examines the feasibility of each solution alternative for review by management. This stage tries to answer the questions, "What do the existing systems do?" "What are their strengths, weakness, trouble spots, and problems?" "What user information requirements must be met by the solution?" "What alternative solution options are feasible?" "What are their costs and benefits?" "Whether the Projects should be executed by external agency?"

Answering these questions requires extensive information gathering and research; careful study of current systems manual or computerized that might be replaced or enhanced as part of this project, seeing through documents, reports, and work papers produced by existing systems; observing how these systems work; polling users with questionnaires; and conducting interviews. In this phase, system analysts work like students to learn and find out users requirements. All of the information gathered during the systems study phase are used to determine information system requirements.

Next, the requirements are studied and structured in accordance with their inter-relationships and eliminate any redundancies. Third, alternative initial design is generated to match the requirements. Then, these alternatives are compared to determine which alternative best meets the requirement in terms of cost and labour to commit to development process. In this phase,

feasibility study of the proposed system is also performed. The main three feasibilities are:

- Technical feasibility which involves question such as whether the technology needed for the information system exists and whether the firm has enough experience using that technology.
- Economic feasibility which involves question such as whether the firm can afford to build the information system, whether its benefits should substantially exceed its costs, and whether the project has higher priority than other projects that might use the same resources.
- Organisational feasibility which involves questions such as whether the information system has enough support to be implemented successfully, whether it brings as excessive amount of change, and whether the organization is changing too rapidly to absorb it. This also involves time feasibility to answer the question such as what is the total time required for the project and whether the organization is ready to devote that much time.

If the information system appears to be feasible from all aspects then this SDLC phase produces a functional specification and a project plan. In case the proposed system is not feasible to develop, it is rejected at this very step. The functional specification is the description of the alternative solution recommended by the analysis team. Once, the recommendation is accepted by the organisation's management, the project teams can begin to make plans to acquire or hire IT resources in terms of hardware, system software, and human resources (if found necessary during the analysis) to build or operate the system proposed. The overall requirements is finally documented as System Requirement Specification (SRS) for further consulting both by users and developers during other phases of SDLC.

After analysis phase is complete, design of the system begins. The design for an information system can be broken down into logical and physical design specifications. The fourth and fifth phases are devoted to design of the new and enhanced system. During design, IS technical team including system analysts convert the description of the recommended alternative solution into logical and then physical system specifications. Design of all aspects of the system from input and output screens to reports, databases, and computer processes are mandatorily done. Design occurs in two phases, viz., logical design and physical design.

10.5.4 Logical Design

Logical design lays out the components of the system and their relationship to each other, as they would appear to users. It specifies what the system solution would do as opposed to how it is actually implemented physically. It describes inputs and outputs, processing functions to be performed, business procedures, data models, and controls. Controls specify standards for acceptable performance and methods for measuring actual performance in relation to these standards.

Logical design is not tied to any specific hardware and systems software platform. Theoretically, the system could be implemented on any hardware and systems software. The idea is to make sure that the system functions as

intended. Logical design concentrates on the business aspects of the system. Various standard charts and diagrams are prepared by the system analysts to illustrate data flows and processing steps for further use. Because the life cycle emphasizes formal specifications, many of the design and documentation tools, such as data flow diagrams, structure charts, or system flowcharts are used.

10.5.5 Physical Design

Physical design is the process of translating the abstract logical model into the specific technical design for the new system. It produces the actual specifications for hardware, software, physical databases, input/output media, manual procedures, and specific controls. Physical design provides the specifications that transform the abstract logical design plan into a functions system of people and machines.

As said above, conversion from logical design to physical design takes place in this phase of SDLC. For example, diagrams are converted to map the origin i.e. starting point to other processing steps, flow, and processing of data in the system. This conversion is done into a structured systems design that can then be broken down into smaller and smaller units known as modules for further conversion to instruction written in a programming language. Various parts of the system are then designed to perform the physical operations necessary to facilitate data capture, processing, and information output. During the physical design, the analyst team decides the programming language in which the computer instructions will be written in, which database system and file structure will be used for the data, the platform that will be used and the network environment under which the system will be run. These decisions finalize the hardware and software plans initiated at the end of the analysis phase. Now, proceedings can be made with respect to acquisition of any new technology not already present in the organization. The final product of the design phase is the physical system specification in a form ready to be turned over to programmers and other system builders for construction. The physical system specifications are turned over to programmers as the first part of the implementation phase.

10.5.6 Implementation

Implementation is the process of putting a system into operation in an organization. During implementation, system specification (output of the system design phase) is turned into working system that is tested and put into use. Implementation includes coding/programing using computer language(s), testing of the developed software system and installation of the system. In fact large organizations may adopt this phase into two – development of the software by coding/programming and testing of the developed system in one phase and in another phase system testing, training and installation of the developed system. The implementation phase being highly technical with most of the activities done by the IT professionals the three steps – coding, testing, and installation have been put under Installation phase for business management point of view. The Installation phase begins with implementation planning for

- developing the software by computer programmers using computer language,

- the process of creating plans for training,
- conversion from old system to newly developed system, and
- testing of the system by the technical team and users.

During coding, programmers write programs that make up the system. Some systems development projects assign programming tasks to specialists whose work consists exclusively of coding programs. Other projects prefer programmers/analysts who both design and code software programs. Since large systems entail many programs with thousands – even hundreds of thousands – of lines of code, programming teams are frequently used. The training plan explains how and when the user will be trained. The conversion plan explains how and when the organization will convert to new business processes. The acceptance-testing plan describes the process and criteria for verifying that the information system works properly in supporting the improved work system. During the system testing, programmers and analysts test the individual programs and the entire system in order to find and correct errors. This is a very crucial and important phase of SDLC because this is the phase which gets the system developed as per users requirements well defined earlier in the analysis phase in terms of SRS. Exhaustive and thorough testing must be conducted to ascertain whether the system produces the right results. Testing answers the question, “Will the system produce the desired results under known conditions?” Testing is also time-consuming. Test data must be carefully prepared, results reviewed, and corrections made in the system. In some instances, parts of the system may have to be redesigned. Yet the risks of glossing over this step are enormous.

Testing information system can be broken down into three types of activities: Unit testing or program testing: it consists of testing each program separately in the system. Testing should be viewed as a means of locating errors in programs, focusing on finding all the ways to make a program fail. Once pinpointed, problems can be corrected.

System testing: it consists of testing the functioning of the information system as a whole. It tries to determine if discrete modules will function together as planned and whether discrepancies exist between the ways the system actually works and the way it was conceived. Among the areas examined are performance times, capacity for the storage and handling peak loads, recovery and restart capabilities, and manual procedures.

Acceptance testing it provides the final certification that the system is ready to be used in a production setting. Systems tests are evaluated by users and reviewed by management. When all parties are satisfied that the new system meets their standards, the system is formally accepted for installation.

It is essential that all aspects of testing be carefully thought out and that they be as comprehensive as possible. To ensure this, the development team works with users to devise a systematic test plan. The test plan includes all of the preparations for the series of tests previously described.

Users play a critical role in the testing process. They understand the full range of data and processing conditions that might occur within their system. Moreover, programmers tend to be aware only of the conditions treated in their programs; the test data they devise are usually too limited. Therefore,

input from other team members and users will help ensure that the range of conditions included in the test data is complete. Users can identify frequent and less common transactions, unusual conditions to anticipate, and most of the common types of errors that might occur when the system is in use. User input is also decisive in verifying the manual procedures for the system.

During installation, conversion takes place. Conversion is the process of changing from the old system to the new system. It answers the question, “Will the new system work under real conditions?” As the result of conversion process, the new system becomes a part of the daily activities of the organization. Application is installed or loaded, on existing or new hardware and users are introduced to new system and trained. In fact, the system analysts begin planning for testing and installation as early as the project initiation and planning phase, since testing and installation require extensive analysis in order to develop the right approach. For conversion, four main strategies can be employed:

- the parallel strategy,
- the direct cutover strategy,
- the pilot study strategy, and
- the phased approach strategy.

In a parallel strategy, both the old system and its potential replacement are run together for a time until everyone is assured that the new one functions correctly. This is the safest conversion approach because, in the event of errors or processing disruptions, the old system can still be used as a backup. However, this approach is very expensive, and additional staff or resources may be required to run the extra system.

The direct cutover strategy replaced the old system entirely with the new system on an appointed day. At first glance, this strategy seems less costly than parallel conversion strategy. However, it is a very risky approach that can potentially be more costly than parallel activities if serious problems with the new system are found. There is no other system to fall back on. Dislocations, disruptions, and the cost of corrections may be enormous.

The pilot study strategy introduces the new system only to a limited area of the organization, such as a single department or operating unit. When this pilot version is complete and working smoothly, it is installed throughout the rest of the organization, either simultaneously or in stages.

The phased approach strategy introduces the new system in stages, either by functions or by organizational units. If, for example, the system is introduced by functions, a new payroll system might begin with hourly workers who are paid weekly, followed six months later by adding salaried employees who are paid monthly to the system. If the system is introduced by organizational units, corporate headquarters might be converted first, followed by outlying operating units four months later.

A formal conversion plan provides a schedule of all the activities required to install the new system. The most time-consuming activity is usually the conversion of data. Data from the old system must be transferred to the new system, either manually or through special conversion software

programs. The converted data then must be carefully verified for accuracy and completeness.

Moving from an old system to a new one requires that end users be trained to use the new system with ongoing user assistance. Detailed documentation including SOPs (Standard Operating Procedures) showing how the system works from both a technical and end-user standpoint is finalized during conversion time for use in training and everyday operations. Documentation is produced throughout the lifecycle. With the adoption of Computer Aided Software Engineering (CASE) tools, more of the documentation is basically a compilation of data and diagrams already stored on a computer. Implementation can continue for as long as the system exists since ongoing user support is also part of implementation. Lack of proper training and documentation contributes to system failure, so this portion of the systems development process is very important. Figure 10.3 illustrates the steps performed in Installation Phase of SDLC.

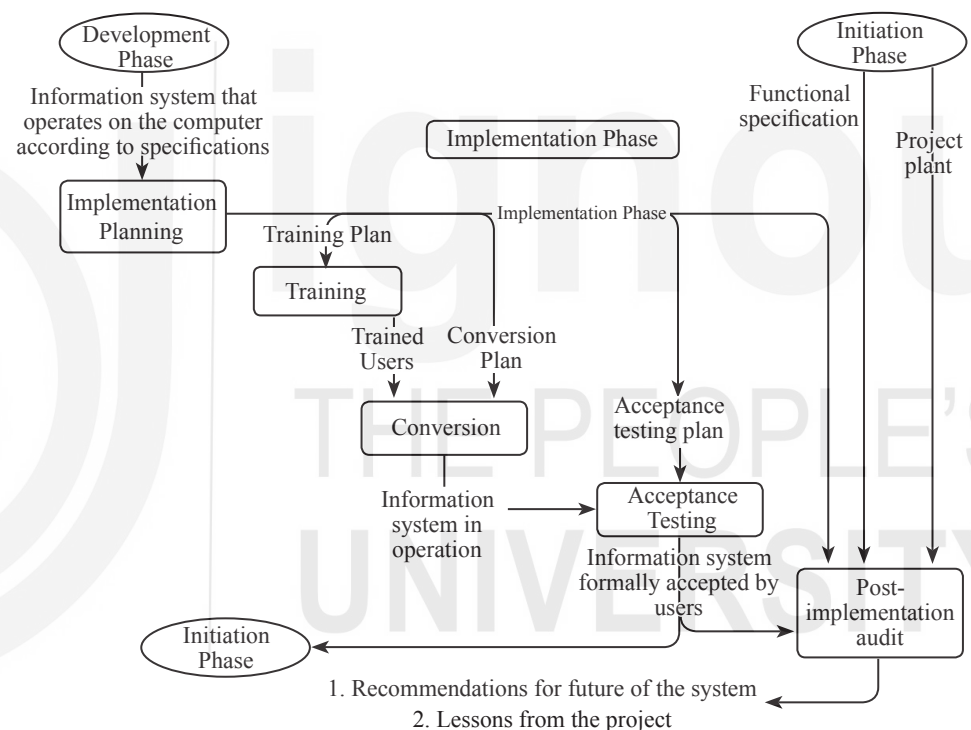


Figure 10.3: Steps in the Implementation Phase of SDLC

10.5.7 Operation and Maintenance

The operation and maintenance phase starts after the users have accepted the new system. This phase can be divided into two activities:

- ongoing operation and support and
- maintenance of the system.

Unlike the other steps in the life cycle, these steps continue throughout the system's useful life. The end of a system's life cycle is its absorption into another system or its termination.

Ongoing operation and support is the process of ensuring that the technical system components continue to operate correctly and that the users use it effectively. It works best when a person or group has direct responsibility for keeping the information system operating. This responsibility is often

split, with the technical staff from the IT Service department taking care of computer operations and members from the users departments in the organization ensuring that users understand the system and use it effectively.

Day-to-day computer operations typically include scheduled events such as generating summary reports from management and backups of the database. The operations manual specifies when these jobs should be done. When the database becomes almost full, or when response times start to increase, the technical configuration of the information system must be changed. This is done by allocating more disk storage space and unloading (by backing up or discarding) data that are not current. With storage space being cheaper as compared to yester years, organizations maintain systems with large storage even on cloud as part of IAAS facility by third party. There are many IT service providers giving Infrastructure-As-A-Service. If the organization opts to go for the complete IS facility by the third party, the total solution may be possible to have as SAAS (Software-As-A-Service). In this case user organization gets just a link to operate using internet facility. Rest is maintained by the IT service provider. In such a case the agreement between both the organization and the IT service provider is signed with mutually agreed terms and conditions.

Maintenance is the process of modifying the information system over time. As users gain experience with a system, they discover its shortcomings and usually suggest improvements. The shortcomings may involve problems unrelated to the information system or may involve ways that the information system might do more to support the work system, regardless of the original intentions. Some shortcomings are bugs. Important shortcomings must be corrected if users are to continue using an information system enthusiastically. Software programers make the changes that users ask for and modify the system to reflect and support changing business conditions. These changes are necessary to keep the system running and useful. Maintenance is not separate phase but a repetition of the other SDLC phases required to study and implement the needed changes. Thus, maintenance is an overlay to the life cycle rather than a separate phase. The amount of time and effort devoted to maintenance depends a great deal on the performance of the previous phase of life cycle. There comes a time when an information system is no longer performing as desired, when maintenance cost becomes prohibitive, or when the organization's needs has changed substantially. Such problems are an indication that it is the time to begin designing the system's replacement, therefore, completing the loop and starting the life cycle over again. Often, the distinction between the major maintenance and new development is not clear, which is another reason why maintenance often resembles the lifecycle itself. Maintenance is of three types:

Corrective maintenance: In this type, the errors that creep into the system are removed, hence the name corrective maintenance.

Adaptive maintenance: It is done to adapt with the changing external factors. For example, if the government rules change regarding the Dearness Allowance from X% to Y%, then the changes have to be made in the Information System to adapt with the changing scenario.

Perfective maintenance: This is done to satisfy the users' requirements to make the system more and more perfect.

The steps in each of the phases of the SDLC have now been introduced. The SDLC is a highly linked set of phases where output of one phase serves as input to the subsequent phase. Throughout the systems development life cycle, the systems development project needs to be carefully planned and managed. Therefore, the larger the project, the greater is the need for project management. Table 10.1 outlines the steps in each phase and makes two major points in addition to the details it presents. First it shows that users are highly involved in most of the phases. In other words, building information systems is not just technical work done by the technical staff. It also shows that each phase has specific deliverables that help maintaining the project progress to date and help keep the project under control. Table 10.1 illustrates the steps and deliverables in SDLC. It helps as a guideline for all - management, IT staff, and users involved in IS system life cycle.

Table 10.1: Step and Deliverables in the Traditional System Life Cycle

Phase/Step	Degree of user participation	Key deliverable, plan or document	Key participants
Initiation			
• Feasibility study	High	Functional specification	User representatives, management, and technical staff
• Project planning	Medium	Project plan	User representatives, management, and technical staff
Development			
• Detailed requirements analysis	High	External specification	User representatives, management, and technical staff
• Internal system design	None	Internal specification	Programmers and technical staff
• Hardware acquisition and installation	None	Hardware plan Hardware operational	Technical staff
• Programming	None	Individual programs debugged	Programmers
• Documentation	Medium	User and programmer documentation	Technical staff and users
• System testing	Medium	Test plan Completed system test	Programmers and users
Implementation			
• Implementation planning	High	Implementation plan	Training, users, and management
• Training	High	Training materials	Trainers and users
• Conversion	High	System in use	Users and project team

• Accepting testing	High	System accepted	Users and project team
• Post implementation audit	High	Audit report	Users and management
Operation and Maintenance			
• Ongoing operation and support	Low	Operations manual	Technical staff
	Low	Usage statistics	Technical staff and users
	High	Enhancement requests and bug fix requests	Technical staff and users
• Maintenance	Medium	Maintenance plan	Technical staff and users
• Absorption or termination	-	-	-

10.8 SUMMARY

There is a fundamental dilemma faced by anyone developing a computer application. Most problems are so large they have to be split into smaller pieces. The difficulty lies in combining the pieces back into a complete solution. Often each piece is assigned to a different team, and sometimes it takes months to complete each section. Without a solid plan and control, the entire system might collapse. Thousands of system development projects have failed or been cancelled because of these complications.

Partly because of the problems that have been encountered in the past, and partly because of technological improvements, several techniques are available to develop computer systems. The most formal approach is known as the systems development life cycle. Any major company that uses SDLC also has to have documents explaining all the steps in various phases including SOPs. Mostly they are softcopies stored for reference as and when required. These documents also contain rules that MIS designers have to follow. Although these details vary from firm to firm, all of the methods have a common foundation. The goal is to build a system by analyzing the business processes and breaking the problem into smaller, more manageable pieces.

10.9 UNIT END EXERCISES

1. What are the different phases of traditional system life cycle?
2. What are the three phases of traditional system life cycle where users are highly involved?
3. What are the various steps and deliverables in the development phase?
4. In which phase of system life cycle the following are performed?
Defining the problem, identifying its causes, specifying the solution, and identifying the information requirements.
5. What are the three major areas of feasibility, which are addressed in system analysis?

6. Which of the design lays out the components of the system and their relationship to each other, as they would appear to users?

10.10 REFERENCES & SUGGESTED FURTHER READING

Alter Steven (1999), *Information Systems (A management perspective)*, Pearson Education

Boyd Donald, W (2001), *Systems Analysis and Modeling (A Macro-to-Micro Approach with Multidisciplinary Applications)* Academic Press and Harcourt India

Ghezzi Carlo, Jazayeri Mehdi and Mandrioli Dino (1991), *Fundamentals of Software Engineering*, Prentice-Hall of India, New Delhi.

Pressman, Roger S. (2001), *Software Engineering (A Practitioner's Approach)*, McGraw-Hill, New Delhi



UNIT 11: COMPUTER PROGRAMMING AND LANGUAGES

Structure

- 11.1 Introduction
- 11.2 Objectives
- 11.3 Programming vocabulary
- 11.4 Control structure or control statement
- 11.5 Overview and features of Visual Basic (VB.Net)
- 11.6 Overview and features of Java
- 11.7 Overview and features of HTML
- 11.8 Overview and features of Excel
- 11.9 Summary
- 11.10 Self-Assessment Exercises
- 11.11 References & Suggested Further Reading

11.1 INTRODUCTION

As we know that to communicate with a computer, we have to write/code in computer language in terms of instructions to the computer. It means that for any problem to be solved by computers, we have to write a program for solving the problem. A computer program is the set of instructions or statements written in any computer programming language. Programming language is the only means by which we can communicate with computer. In other words computer programs work as an interface between a person and a computer. There are many computer languages to write programs. Broadly they are categorized as Low Level language e.g. Machine Level; Assembly Level language; and High Level language (e.g. C, C++, Visual Basic (VB.NET), Python, JAVA etc.). Low-level language is hard to learn as computer programs are required to be coded in '0' and '1'. Similarly Assembly level is also not easy as programs are required to be written in mnemonic codes. Both these languages are machine dependent and they are used by expert programmers. High-level programming languages are easy to understand and learn as they are written in English like language.

11.2 OBJECTIVES

After reading this unit, you should be able to

- Explain the vocabulary used in programming languages;
- Identify and describe control structures in programming languages;
- Identify important features of VB.NET, Java, HTML and XML, Excel, and Python;
- Write programs in higher level language.

11.3 PROGRAMMING VOCABULARY

Suppose we want to write program for finding factorial of a number. The flow chart and pseudo code for this will be as follows:

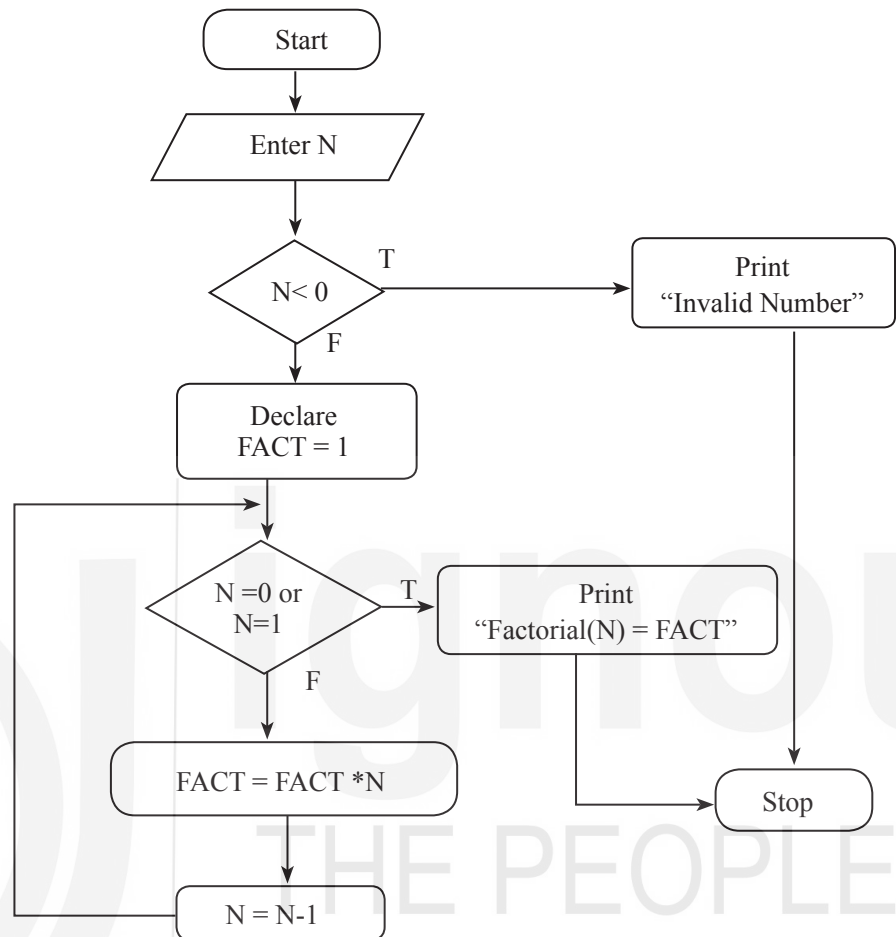


Figure 11.1(a)

Pseudo code for the above is as given below:

```
Step 1 :   Input N
Step 2 :   if (N>0)
Step 3 :       Declare FACT = 1
Step 4 :       if((N=0) or (N=1))
Step 5 :           Print "Factorial(N) = FACT"
Step 6 :       else
Step 7 :           Repeat
Step 8 :               FACT = FACT*N
Step 9 :               N=N-1
Step10 :          Until(N>0)
Step11 :   else
Step12 :       Print "Invalid Number"
Step13 :   end
```

Figure 11.1(b)

In writing above program we have used some important words such as ‘if – then –else’, ‘repeat – until’, Input, Print, Declare etc. Respective meanings, inherent in these words, form fundamental concepts in programming. These words are called ‘Reserve Words’ or ‘Keywords’ of a programming language. To write program in any language is simply a matter of learning its syntax and semantic rules and special features. While syntax denotes the grammar rules, semantic denotes the meaning associated. ‘N’ denotes the number for which factorial is required to be obtained by the computer. ‘FACT’ denotes the calculated factorial of the number N during the execution of the program and till the end.

Apart from key words a program uses several other programming concept viz. identifiers, constants, expression, library functions etc. These concepts are defined here below.

Identifiers:

Identifiers are the name given to various program elements such as variables, functions, sub-routines etc. Identifiers consist of letters and digits, in any order, except the first character must be a letter. For example, the name FACT used in figures 11.1(a) and 11.1(b) is an identifier. It has been used to denote/identify the values of factorial of ‘N’ which is a number to be provided as input.

Variables:

Variables are those elements of program whose value can change during the execution of a program. Both ‘N’ and ‘FACT’ used in the figures referred above are variables as their values are changing after each computation.

Constant:

Constant are those elements of program whose value cannot be changed during the execution of the program. Means they have fixed value.

The numeric constant can be integer representing whole number or floating point number with a decimal point. Constants would be probably the most familiar concept to us since we have used them in doing everything that has to do with numbers. Operations like, addition, subtraction, division, multiplication, and comparison can be performed on numeric constants.

String constants are sequence of alphanumeric characters i.e. combination of ‘alphabets’ and ‘numerics’ enclosed in double quotation marks e.g. “Computer”, “X123”. String constant can be concatenated and compared in a lexicographic sense.

Expression:

An expression represents a single data item, such as a number or a character. The expression may consist of a single entry such as a constant, a variable, an array, or a reference to a function. It may also consist of some combination of such entities inter connected by one or more operators. For example $FACT = 1$, $FACT = FACT * N$ etc. are expressions.

Generally every high level programming language provides some sets of arithmetical, relational, logical operators. For knowledge about computer programs you should be aware about theses basics. The most common arithmetical, relational, and logical operators, as we all know, are:

Arithmetical operators:

* for Multiplications

/ for Divisions

- for Subtractions

+ for Additions

Relational operators

= Equal to

<> Not equal to

< Less than

> Greater than

<= Less than or equal to

>= Greater than or equal to

Logical Operator

NOT Logical NOT operator simply negates a truth-value.

AND Logical AND operator return true if both operators are true (e.g. $(4 \leq 8)$ and $(a \leq b)$ is true.

OR Logical OR operator return false if both operands and false.

XOR Exclusive OR is true only if one of the operand is true and other is false.

Each operator has got some precedence associated with it and expression is always evaluated on the basis of precedence of the operator e.g.

$3 + 7 * 6 - 4 / 2 - 2$ 4

In above expression exponential operator is evaluated first then multiplication and division operators are evaluated and lastly the addition and subtraction operator are evaluated. The precedence of arithmetic operators is exactly the same as in mathematics. That is computer follows BODMAS (Bracket Of, Division, Multiplication, Addition, Subtraction). If the expression contains Exponential operator then it is of highest precedence i.e. it is evaluated first.

Relational operators have same level of precedence among themselves but low precedence than arithmetic operators. Logical operators have lower precedence than relation operators. Among themselves they have the following order of priority during evaluation of logical expression (logical expression is an expression evolving relational or logical operators and evaluate either true value or false value)

NOT Highest precedence

AND

OR, XOR Lowest precedence

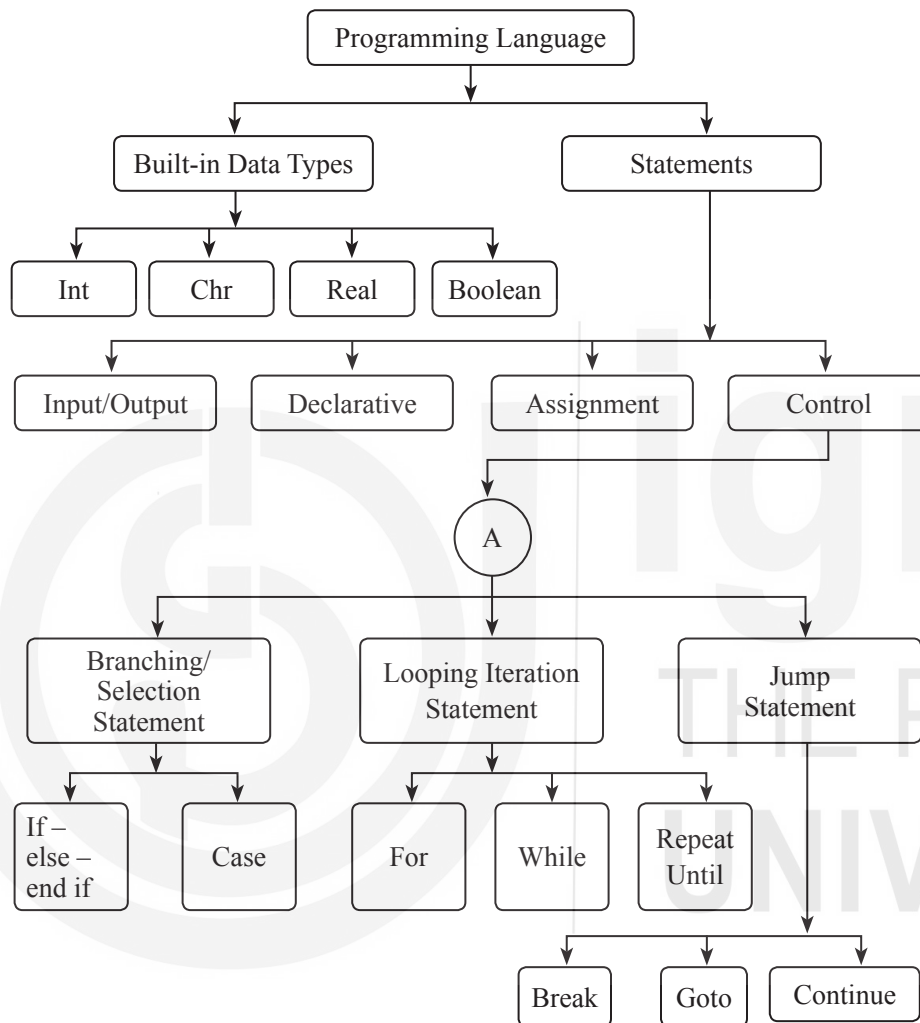
Expressions would be our basic workhorses. Our programs would be full of them, and while executing the program; computer would be preoccupied most of the time to evaluate them. But it would also be doing other things such as taking some values as input from input device, assigning a value or result of an expression to a variable, send a value for display to output device

pondering to be or not to be on some conditional expression, repeating some tasks until we are satisfied and calling some slave – program to perform specific task. In the following discussion we shall elaborate on these actions.

Generally every high level programming language has got some

- i) Built-in Data Types
- ii) Statements

They are illustrated here below:



Built-in data type and statement type are specific to a particular programming language.

Built in Data Types

Generally all programming language support following types of built in data viz.

Int.	Whole number
Real	Number with fractional part
Boolean	Data type which is either 'true' or 'false'
Char	Single alphabet or numeric character enclosed in single quotation mark e.g. 'a', '1' etc.
String	Group of characters enclosed in double quotation mark.

Before using any variable in the program it is generally necessary to define its type, broadly known as 'data type'. The computer memory requirement for each data type is different and it is dependent upon the specific programming language.

Statements

Statement in programming language can be broadly classified as follows:

- i) Declarative statements
- ii) Input output statement
- iii) Assignment statement
- iv) Control statement or control structure.

Declarative Statements

These statements are used to declare the type of the variable e.g.

Examples in Java -

Int x : x is of integer type variable
Char x : x is character type variable
Byte b : b is of byte type variable
Float y : y is of float (real) type variable

Examples in Visual Basic -

Dim x as Integer : x is integer type variable
Dim b as Byte : b is Byte type variable
Dim y as Single : y is of floating (real) type variable

Input Output Statement

They can be broadly classified as

- i) Console i.e. on screen input statement
- ii) File input statement for giving input in a file
- iii) Console i.e. on screen output statement
- iv) File output statement for getting output in a file

Console input is used for taking the input from default input devices e.g. Step 1 of Figure 11.1(b) while console output statement is used for printing the result of the program on default output device e.g. Step 4, and Step 12 of Figure 11.1(b). Similarly file input statements are used for taking the input from file or non default input device while file output statement is used for writing the output into the specified file.

Assignment statements are used for assigning the value to variable e.g. variable name = value to be assigned as in Step 3 and Step 9 of above referenced figures. Some more examples may be as given below -

X = 123 : value 123 is assigned to X
Z = func1() : return value of func1() will be assigned to the variable Z. Here data type of Z and data type of value returned by func1() must be of the same type.

11.4 CONTROL STATEMENT OR CONTROL STRUCTURE

The execution of computer program is always sequential i.e. statement which is written first will be executed first. To change the sequential execution of a program, we need control structures which can be broadly categorized as follows:

- i) Branching or selection statement
- ii) Looping or Iterative statement
- iii) Jump statement

Branching or selection statements

They are used to select one group of statements or single statement for execution from several available groups of statements. The selection of a particular group depends on some Boolean condition. Example -

If – then – else

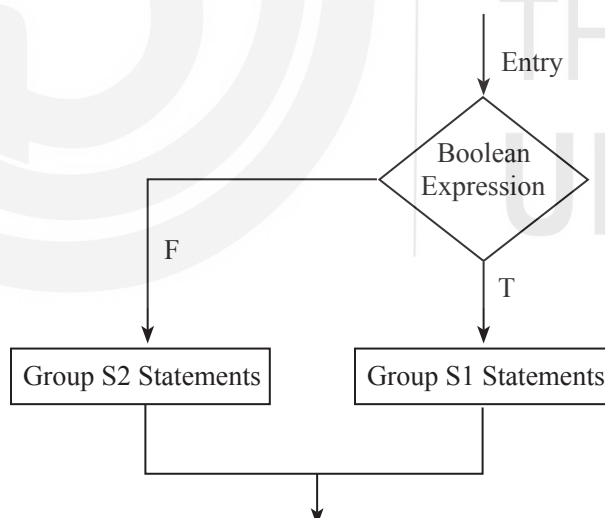
This statement is used when we have to make selection from two available groups of statements. S1 group of statements will be only executed if Boolean expression is true otherwise S2 group of statement will be executed.

If <Boolean expression>

S1 statement(s)

Else

S2 statement(s)



CASE statement is used if we have to make selection from several available groups of statements

Case (Expression R)

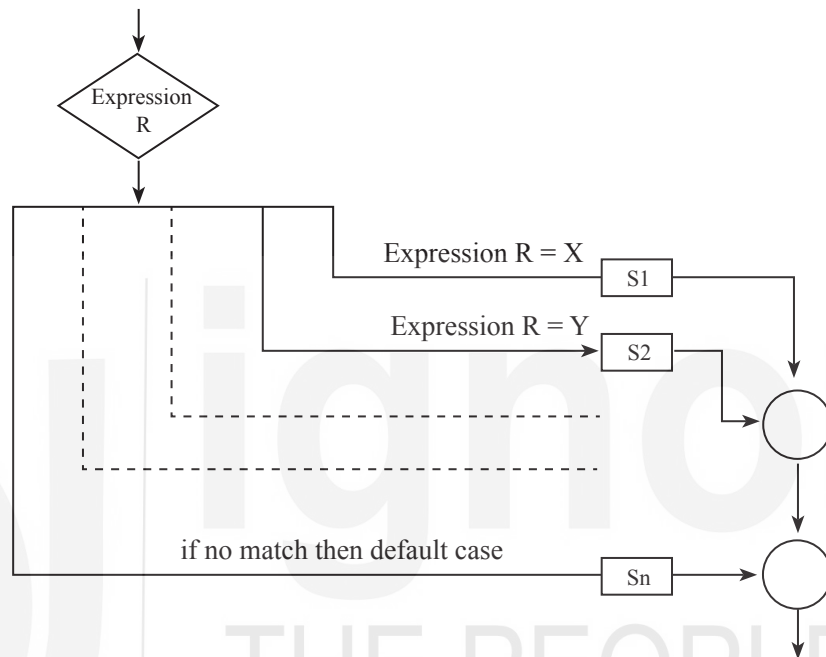
{	Case X	:	S1
	Case Y	:	S2
	Case W	:	S3
	Case Z	:	S4

.....
.....

default : Sn (default statement)

}

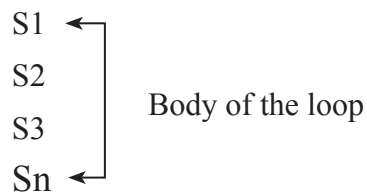
Here X, Y, W, Z are permissible values for Expression R. If Expression R evaluates to value Y then S2 group of statements will be executed if Expression R evaluates to non-permissible values i.e. none of X, Y, W, and Z then default i.e. Sn group of statements will be executed.



Looping / Iterative Statement

A computer is well suited to perform repetitive operations. It can do it tirelessly. Every computer language must have features that instruct a computer to perform such repetitive tasks. The process of repeatedly executing a block of statements is known as looping. The statements in the block may be executed any number of times, from zero to infinite number. If a loop continues forever it is called an infinite loop. Finite loop is a loop, which terminates after fixed number of iterations. These fixed number of iterations are known to us in advance for implementing finite looping we use for loop structure.

For Variable X = value1 to value2 steps size value3

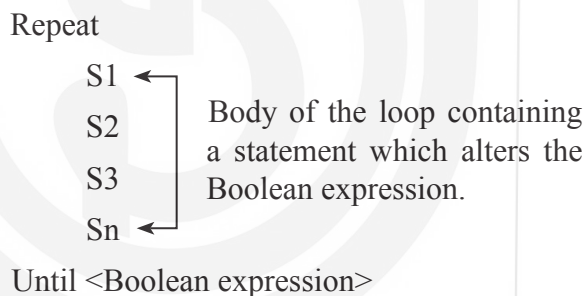
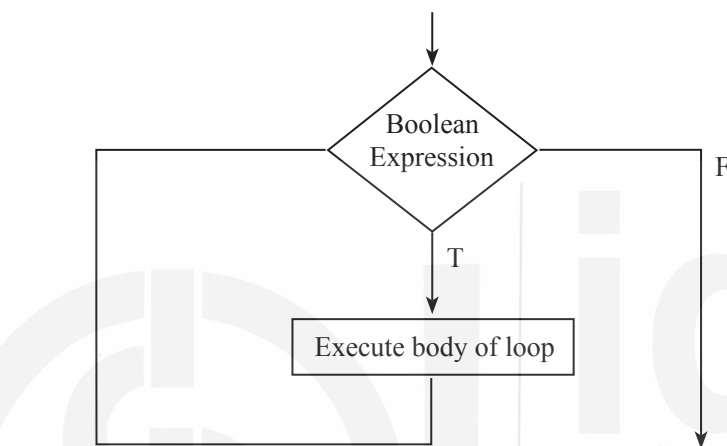
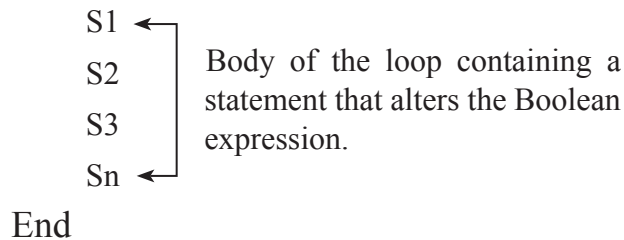


End

In this case Variable X is initialized to value1, then it checks whether the value of variable is greater value2. If not then it executes the body of the loop and then increase the value of variable by step size equal to value 3.

If we do not know in advance how many times the loop will iterate then we use while – loop or repeat – until structures. There types of loops are terminated by some Boolean expression. The body of these loops must contain one statement that alters the value of the Boolean expression.

While <Boolean expression>



It is clear from the flow chart of ‘while’ loop and ‘repeat – until’ loop that body of repeat until loop will always executed at least for one time.

Jump Statement

Goto, break and continue are generally the jump statements. Goto statement is used for jumping from one location to another labeled location within the program. Break statement causes the loop to be terminated, the continue statement causes the loop to be continued with the next iteration after skipping any statements in between.

Functions/ Sub-routine / Sub program/ Modules

There are two ways to write a large and complicated computer program. One is to write the single program, if we can, which are normally very difficult for a large and complicated program. The other is to look at the large program as consisting of several somewhat less complicated subprograms. Coping with each of the subprogram can be seen in the same manner as main program, either write them directly, if we can, or refine it in terms of further subprograms and so on. The process of breaking large

and complicated programs in smaller and less complicated sub program is known as modular programming approach. Each subprogram is termed as function, subroutine, module etc. depending on the specific programming language used to write the program.

11.5 OVERVIEW AND FEATURES OF VISUAL BASIC (VB.NET)

Visual Basic, originally called **Visual Basic .NET (VB.NET)**, was launched by Microsoft in 2002 as the successor to its original Visual Basic language, the last version of which was Visual Basic 6.0 (VB6). The “.NET” portion of the name was dropped in 2005.

“.NET” (pronounced as dot net) is a framework that provides a programming guidelines that can be used to develop a wide range of general purpose applications, from web to mobile to Windows based applications. The applications developed using VB.NET can also be run in other systems having Linux, Mac operating systems.

Visual Basic.NET (VB.NET or VB.Net) is an object-oriented computer programming language implemented on the .NET Framework. Although it is an evolution of classic Visual Basic language, it is not backwards-compatible with VB6, and any code written in the old version does not compile under VB.NET. It is easy to learn and has strong programming features. Object-oriented programming (OOP) is a programming paradigm based on the concept of “objects”, which can contain data and code: data in the form of fields (often known as attributes or *properties*), and code, in the form of procedures (often known as *methods*).

The following reasons make VB.Net a widely used professional language –

- It is one of the popular modern languages used worldwide for general purpose applications.
- It is based on object oriented concept of modern programming technique.
- It is easy to learn because it's foundation is based on the BASIC (Beginner's All-purpose Symbolic Instruction Code) programming language and VB6.
- It is a structured language. The Structure data type is user defined and can be used to represent information about something more complicated than a single number (Integer), Character, or Boolean data types can do.
- It produces efficient programs meaning thereby that the codes written in VB.NET use optimal processing time and computer memory space.
- It can be compiled on a variety of computer platforms like Windows, Linux, Mac, and mobiles.

In computer programming, Data types refer to an extensive system used for declaring variables or functions of different types. The type of a variable determines how much space it occupies in storage and how the bit pattern stored is interpreted.

Data Types Available in VB.NET

VB.Net provides a wide range of data types. The following table shows all the data types available –

Table 11.1: Data types in VB.Net

Data Type	Storage Requirement
Boolean	Depends on implementing platform
Byte	1 byte
Char	2 bytes
Date	8 bytes
Decimal	16 bytes
Double	8 bytes
Integer	4 bytes
Long	8 bytes
Object	4 bytes on 32 bit platform 8 bytes on 64-bit platform
SByte	1 byte
Short	2 bytes
Single	4 bytes
String	Depends on implementing platform
UInteger	4 bytes
ULong	8 bytes
User-Defined	Depends on implementing platform
UShort	2 bytes

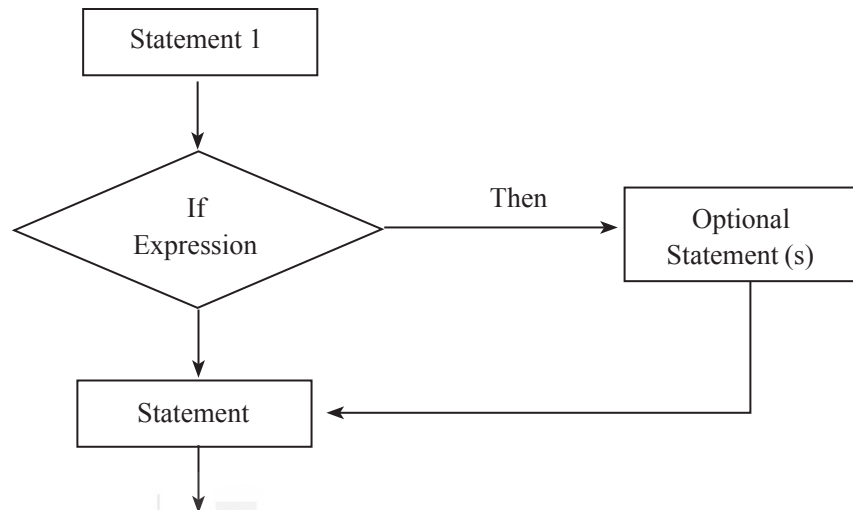
Control Statements of VB.NET

Following types of conditional branching statements are available in Visual Basic

- i) If – then – end
- ii) If – then – else
- iii) If – then – Else if - End if
- iv) Select case statements

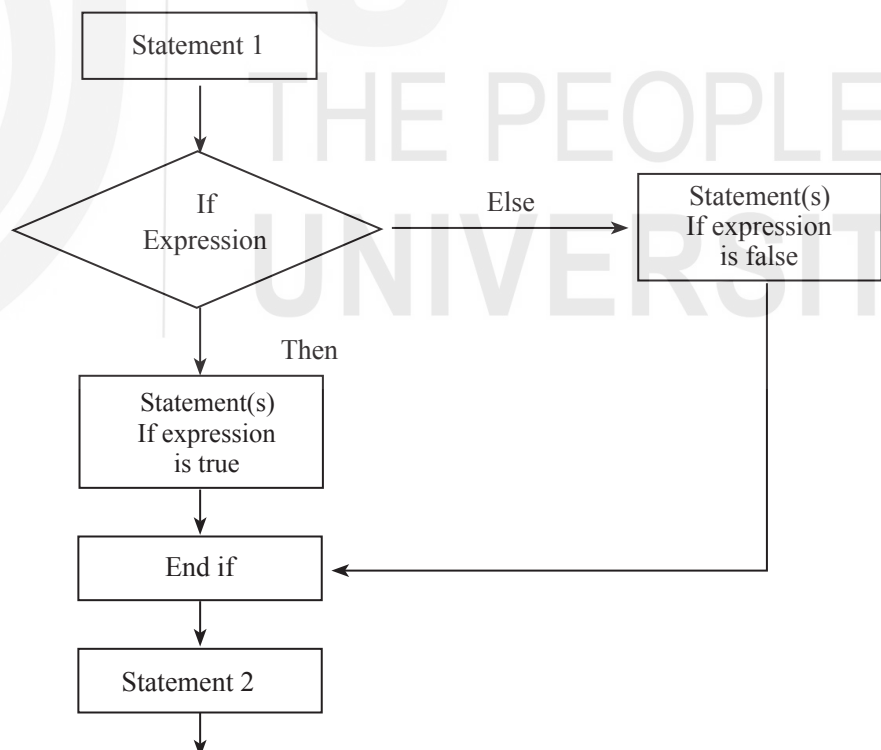
If- then is most common conditional branching statement. This is simplest and illustrated below. Syntax of If- then – end is

If *expression* then
Statement(s)
End if



The optional statement block is executed only when expression is true. Regardless of the value of the expression, execution continues after the ‘if’ statement with statements.

If-then else is used when certain statements execute only when the expression is true other statements execute if the expression is false



Syntax:

If *Expression* Then
Statements(s)

End if

```
graph TD; S1[Statement 1] --> D1{If Expression 1}; D1 -- Then --> S1T[Statement(s) if expression -1 is true]; D1 -- Else --> D2{If Expression 2}; D2 -- Then --> S2T[Statement(s) if expression - 2 is true]; D2 -- Else --> D3{If Expression 3}; D3 -- Then --> S3T[Statement(s) if expression -n is true]; S1T --> E1[End if]; S2T --> E1; S3T --> E1; E1 --> S2[Statement 2]; S2 --> Exit(( ));
```

If *expression* *I* then
Statement(s)

```

Else if expression 2 then
    Statement(s)

```

:

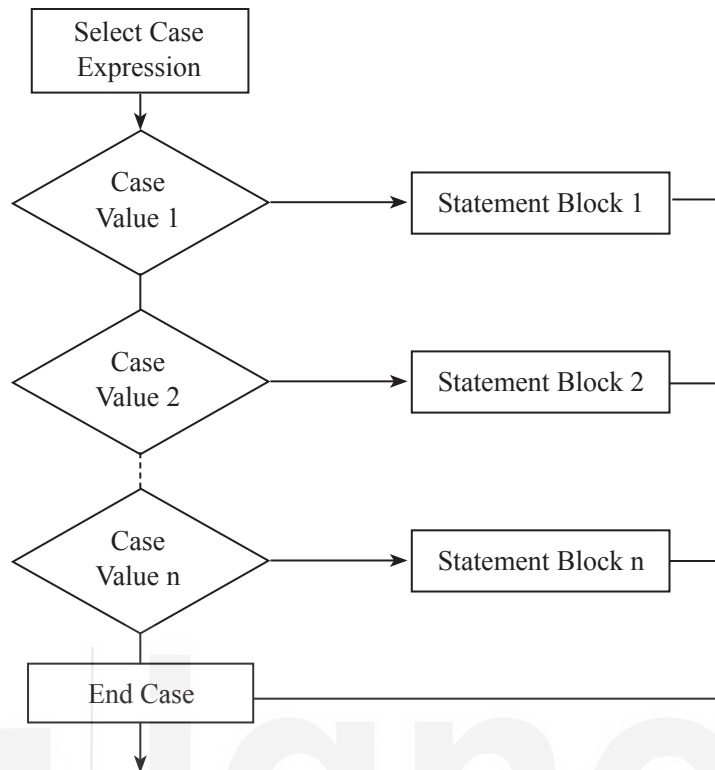
•

•

Else if *expression n* then
Statement(s)

End if

199



Syntax:

```

Select Case Expression
    Case value 1
        Statement(s)
    Case value 2
        Statement(s)
    :
    :
    :
    Case value n
        Statement(s)
Else Case
    Statement(s)
End Select
  
```

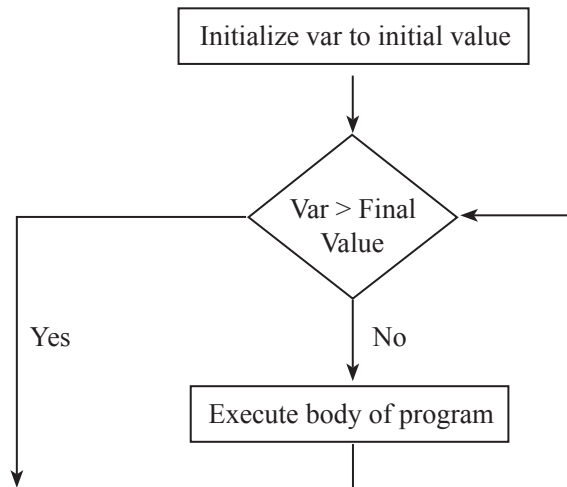
Looping Constructs of Visual Basic

The For... Next is useful whenever we know how many times loop through the statement block

Syntax

```

For var = initial value to final value
    Statement 1
    :
    :
    Statement n
  
```

Do – Loop executes a block of statements for as long as a condition is true. There are two variations of the Do – Loop statements, but both use the same basic model.

The loop can be executed either while condition is true or until the condition becomes true. The two variations of the Do – loop use the keywords while until to specify how long the statements are executed. To execute a block of statements while a condition is true, use the following syntax

```

Do while condition
    Statement block
Loop
  
```

To execute a block of statements until the condition becomes true use the following syntax:

```

Do Until Condition
    Statement block
Loop
  
```

11.6 OVERVIEW AND FEATURES OF JAVA

Java is a high-level, general purpose, object – oriented language originally developed by Sun Microsystems of USA in 1991. It was released in 1995 as a core component of Sun Microsystems' Java-Platform. Sun Microsystems was acquired by Oracle in 2010. As of October 2021, Java 17 is the latest version.

The most striking feature of the language is that it is platform neutral language. It is intended to let programmers *Write Once Run Anywhere* (WORA) meaning that compiled Java code can run on all platforms that support Java without the need for recompilation. Java is the first programming language that is not tied to any particular hardware or operating system. As such, programs developed in Java can be executed anywhere on any system. More than a billion computers and at least 3 billions mobile phones use Java for application development.

Java has become a popular and useful programming language because of its excellent features. The important feature of Java is as under:

- Simple and Familiar
- Compiled and Interpreted
- Platform Independent, Portable, and Architectural Neutral
- Object-Oriented, secure, and Robust
- Distributed, Multi-threaded and Interactive
- High Performance, Dynamic and Extensible

Programs written in Java are simple and easy to understand. It has a base of familiar and pre-existing languages like C and C++ and removes the drawbacks, and complexities of C/C++.

Usually, computer language is either compiled or interpreted. Java combines both these approaches thus making Java a two-stage system. Java compiler translates source code into what is known as byte code instruction. Byte codes are not machine instructions and therefore, Java Virtual Machine (JVM) then executes this bytecode which is executable on many operating systems and is portable. JVM works as the interpreter generating machine code that can be directly executed by the machine that is running Java program. The Figure 11.2 shows the above process:

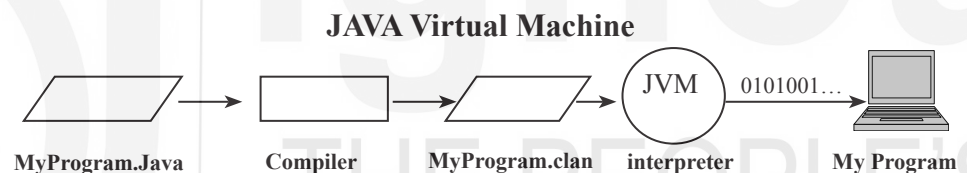


Figure 11.2

The most significant feature of Java is that it provides platform independence which leads to a facility of portability. Being platform-independent means a program compiled on one machine can be executed on any machine in the world without any change. Java achieves platform independence by using the concept of the **BYTE code**. This is where the “*Write Once, run anywhere*” (*WORA*) slogan for Java comes in.

Below diagram explains the platform independence feature of Java-

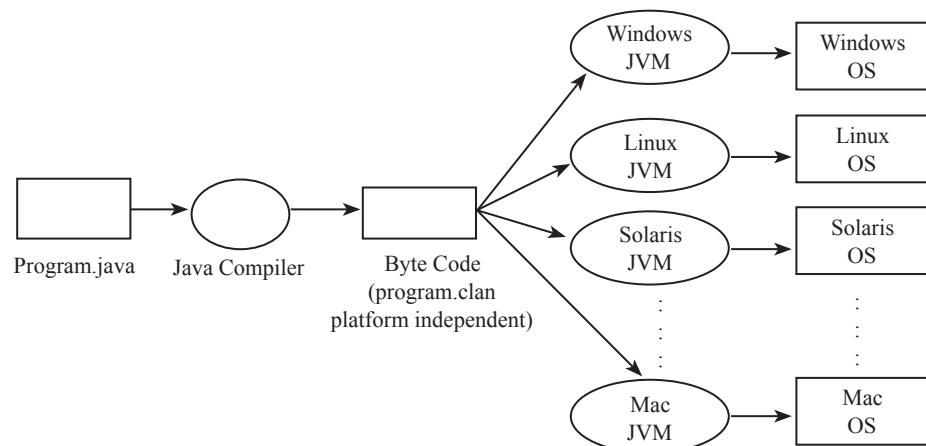


Figure 11.3

Java is a true object oriented language. Almost everything in Java is an object. All program code and data reside within objects and classes. You cannot develop an executable program in Java without making use of the class. This indicates that Java very strictly applies the principle of Encapsulation which is one of the prime features in Object Oriented Programming (OOP).

Security is an important issue for any programming language as there is a threat of malicious activities and viruses. Java supports facilities to check memory access and also ensures that no viruses enter into the code. It also has a bytecode verifier that checks the code fragments for any illegal code that violates the access right.

Java is robust as it is capable of handling run-time errors, supports other features and has a strong memory management system. It helps in eliminating errors as it checks the code during both compile and runtime.

Java is *garbage-collected language* as JVM automatically de-allocates the memory blocks and programmers do not have to worry about deleting the memory manually as in other languages.

Java also provides the concept of exception handling which identifies runtime errors and eliminates them.

Java is designed as a distributed language for creating applications on networks. It has the ability to share both data and programs. Java applications can open and access remote objects on Internet as easily as they can do in a local system. This enables multiple programmers at multiple locations to collaborate and work together on a single project. It encourages users to create distributed applications. You can split a program into many parts and store these parts on different computers. A Java programmer sitting on a machine can access another program running on the other machine. This feature in Java gives the advantage of distributed programming, which is very helpful when we develop large projects.

A thread is an independent path of execution within a program, executing concurrently. Multithreaded means handling multiple tasks simultaneously. Java supports multithreaded programs. This means that we need not wait for the application to finish one task before beginning another e.g. we can listen to an audio clip while scrolling a page and at the same time download an applet from a distant computer.

The performance of Java is impressive for an interpreted language because of its intermediate bytecode. Java provides high performance with the use of “JIT – Just In Time compiler”, in which the compiler compiles the code on-demand basis, that is, it compiles only that method which is being called. This saves time and makes it more efficient. Java architecture is also designed in such a way that it reduces overheads during runtime. The inclusion of multithreading enhances the overall execution speed of Java programs. Bytecodes generated by the Java compiler are highly optimized, so Java Virtual Machine can execute them much faster. Java is dynamic and extensible means that you can add classes and add new methods to classes, creating new classes through subclasses. Java gives the facility of dynamically linking new class libraries, methods, and objects. It is highly dynamic as it can adapt to its evolving environment.

Java Environment

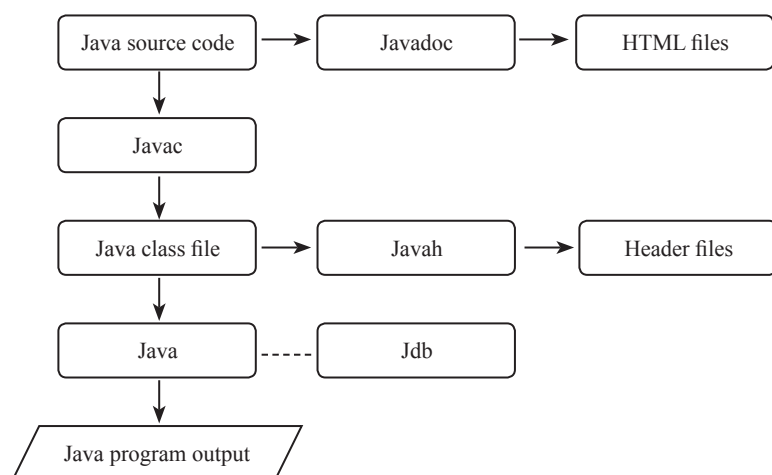
JAVA software for your computer, or the Java Runtime Environment (JRE) is also referred to as the Java Runtime, Java Virtual Machine(JVM), Java Add-on, Java Plug-in etc. JRE is a software layer that runs on top of a computer's operating system software and provides the class libraries and other resources that a specific Java program needs to run. It includes a large number of development tools and hundreds of classes and methods. The development tools are part of Java Development Kit (JDK) and classes and methods are part of Java Standard Library (JSL).

Following Java Basic Tools are the foundation of the JDK. They are the tools you use to create and build applications.

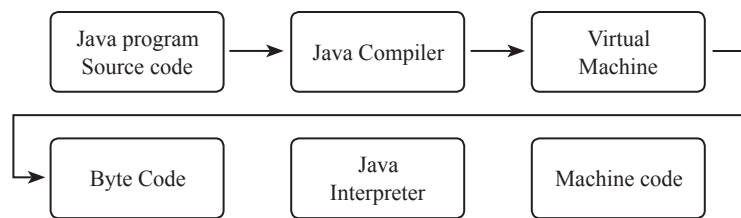
Table 11.2: Java Development Tools

Tools	Description
Appletviewer	To run and debug Java applets without a web browser.
Apt	Annotation processing tool.
Extcheck	Utility to detect Jar conflicts.
Jar	To create and manage Java Archive (JAR) files.
Java	Java interpreter, which runs applets and application by reading and interpreting bytecodes. It is used for both – development and deployment
Javac	The Java compiler, which translates Java source code to byte code files that the interpreter can understand.
Javadoc	API documentation generator. Creates HTML format documentation from Java source code file.
Javah	Produces header file for use with native methods.
Javap	Java disassembler, which enables us to convert bytecode files into a program description.
Jdb	Java debugger, which helps us to find errors in our programs.

The way these tools are applied to build and run application program is show below in the figure



Java standard library includes hundred of classes and methods grouped into six functional packages.



1. Language support package: Collection of classes and methods required for implementing basic features of Java.
2. Utilities package: A collection of classes to provide utility functions such as time and date.
3. Input / output package: A collection of classes required for input/output manipulation.
4. Networking package: A collection of classes for communicating with other computer via intranet or Internet.
5. Awt Package: The abstract window tool kit package contains classes that implements platform independent graphical user interface (GUI).
6. Applet package: This includes a set of classes that allow us to create java applets. Applets are small java programs developed for intranet/ Internet applications.

Some of the examples of Java tools are StdIn.java, StdOut.java, StdAudio.java, StdStats.java, StdRandom.java, In.java, Out.java, Draw.java, Picture.java which are used for their respective applications.

An applet located on a distant computer (server) can be downloaded via internet/ intranet and executed on a local computer (client) using a java capable browser.

Structure of Java program

A typical structure of a Java program contains the following sections:

Sections	Respective uses...
Documentation section	The <i>documentation</i> section is an important section but optional for a Java program. It includes basic information about a Java program. The information includes the author's name, date of creation etc. Whatever one writes in the documentation section, the Java compiler ignores the statements during the execution of the program.
Package statement	The <i>package</i> declaration is optional. It is placed just after the documentation section. In this section, package name in which the class is placed, is declared.
Import statement	The package contains the many predefined classes and interfaces. If we want to use any class of a particular package, we need to import that class. The import statement represents the class stored in the other package. <i>Import</i> keyword is used to import the class.
Interface statement	It is an optional section. You can create an <i>interface</i> in this section if required. 'Interface' keyword is used to create an interface. An interface is a slightly different from the class. It contains only constants and method declarations.

Class definitions	In this section, you define the <i>class</i> . It is vital part of a Java program. Without the class, you cannot create any Java program.
Class Variables and Variables	In this section, you define <i>variables</i> and <i>constants</i> that are to be used later in the program. In a Java program, the variables and constants are defined just after the class definition.
Main method class	In this section, you define the <i>main() method</i> . It is essential for all Java programs because the execution of all Java programs starts from the main() method.
Methods and Behaviors	In this section, we define the functionality of the program by using the <i>methods</i> . The methods are the set of instructions that we want to perform. These instructions execute at runtime and perform the specified task.

All Java source files will have the extension Java, also if a program contains multiple classes, the file name must be the class name of the class containing the main method.

Data Types in JAVA

Data types specify the different sizes and values that can be stored in the variable. There are two types of data types in Java:

1. **Primitive data types:** The primitive data types include boolean, char, byte, short, int, long, float and double.
2. **Non-primitive data types:** The non-primitive data types include Classes, Interfaces, and Arrays.

Table 11.3: Data types in JAVA

Data Type	Storage Requirement
Boolean	1 bit
Char	2 bytes
Byte	1 byte
Short	2 bytes
Int	4 bytes
Long	8 bytes
Float	4 bytes
Double	8 byte

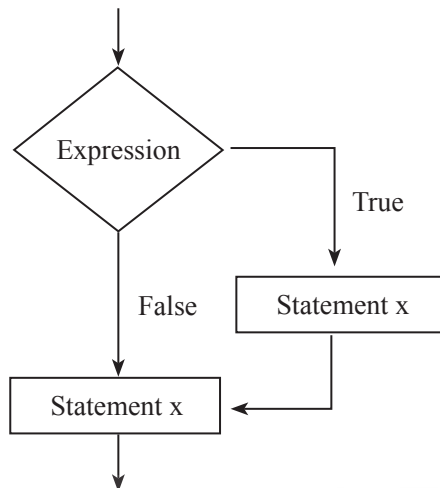
Control Statements

The if-else statement in java may be implemented in following different form depending on the complexity if condition to be tested:

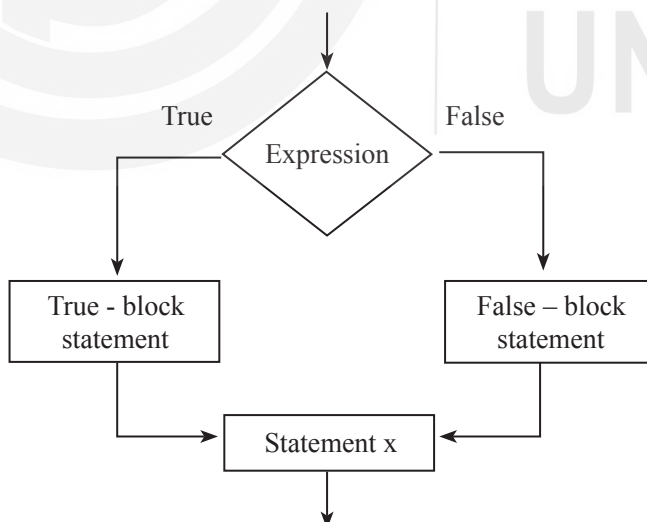
- i) Simple if statement
- ii) If-else statement
- iii) Nested if-else statement
- iv) Else if ladder

```
If(expression)  
{
```

Statement(s)
}
Statement x



If(expression)
{
 True – block statement(s)
}
else
{
 False – block statement(s)
}
Statement – x



If(Condition-1)
 Statement – 1;
Else if (Condition – 2)
 Statement – 2;

Else if (*Condition – 3*)

Statement – 3;

.....

.....

Else if (*Condition – n*)

Statement – n;

Else

Default statements;

Statement – x

Activity A

1. What are the different types of conditional branching statement in Visual basic?
2. Describe the six functional packages in JAVA program library.

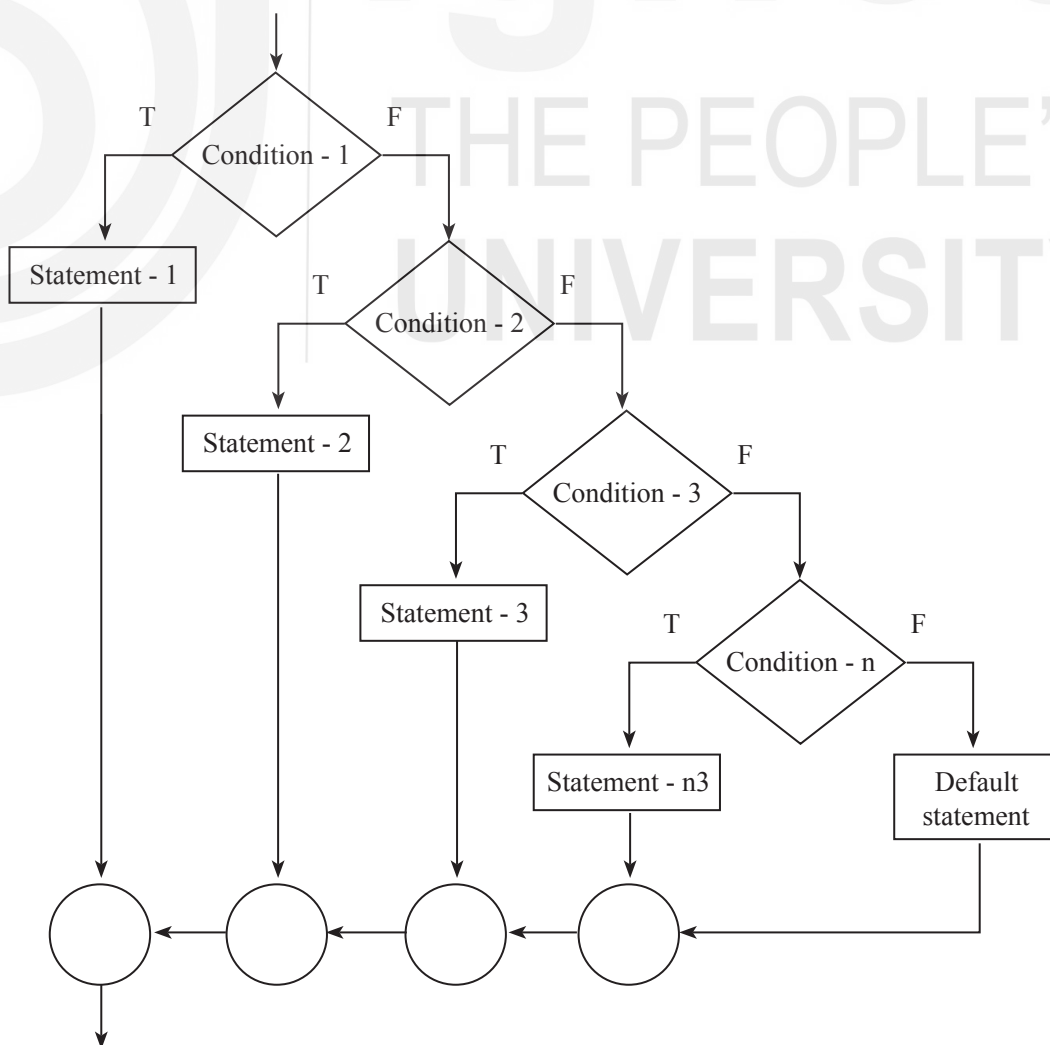
.....

.....

.....

.....

.....



The Java language provides three constructs for performing loop operations
viz

While loop

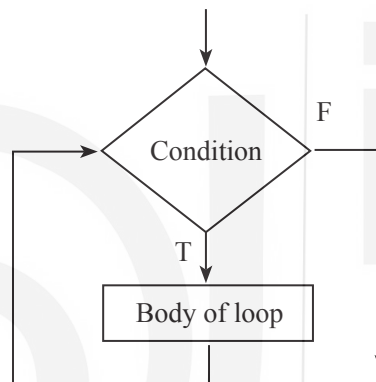
Do loop

For loop

In the while loop test condition is evaluated and if the condition is true, then
body of loop is executed as illustrated in the figure below:

Syntax:

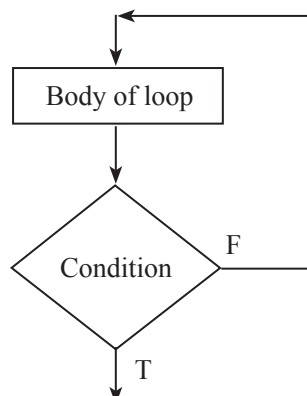
```
Initialization  
While(Condition)  
{  
    Body of loop  
}
```



In do statement test condition is evaluated the bottom of the loop hence
resulting at least one time execution of the body of the loop irrespective of
the value of test condition:

Syntax:

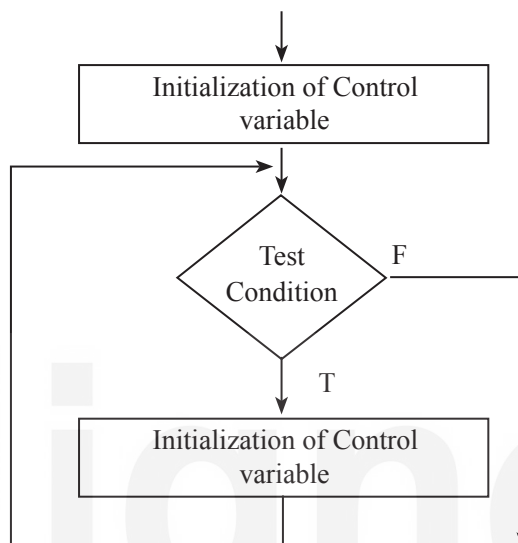
```
Initialization  
Do  
{  
    Body of loop;  
} While (condition)
```



The *for* loop is used when we know in advance how many times the loop will run. The syntax is:

For(initialization; test condition; increment)

```
{  
    Body of loop;  
}
```



11.7 OVERVIEW AND FEATURES OF HYPERTEXT MARKUP LANGUAGE

For web site development and to provide information through web, we prepare documents in a markup language known as Hypertext Markup Language (HTML). It was first created in 1989. As of now, in 2021, the latest version is known as HTML5. HTML is used to create pages and make them functional. Hypertext means that the document contains links that allow the reader to jump to other places in the document or to another document altogether. A Markup Language is a way that computers speak to each other to control how text is processed and presented. To do this HTML uses two things: tags and attributes. With HTML, we can describe the structure of our web documents, including related information such as the integration of multimedia and use of hyperlinks. It also links java applets to our web site.

HTML is actually a subset of an internationally known standard called Standard Generalized Markup Language (SGML). The advantage of HTML is that we can create its documents with simple ASCII text editor, which is not always true for document based on SGML. HTML defines documents so that any browser running on any computer can read a display them.

The basic HTML commands that are needed to create web pages fall into following categories:

Structural Command

These identify a file as an HTML document and provide information about the data in the HTML file.

Paragraph formatting Command

These specify paragraph endpoints and heading levels.

Character formatting Command

These allow us to apply various styles to the characters in our documents.

List specification Command

HTML supports several list formats including bulleted numbered and definition list.

Hyper linking command

These allow us to provide information about moving from one document to another.

Assesst interaction command

These allow us access to multimedia information. Through these commands, we can display graphical images, access sound and provide digital movies for users.

The body of any HTML document is as under:

```
<HTML>
<HEAD>
<TITLE> This is document title </TITLE>
</HEAD>
<BODY>
:
:
:
This is document text
:
:
<BODY>
</HTML>
```

HTML file has .HTM or .HTML as their file name extension.

XML and XHTML

XML (eXtensible Markup Language) is a mark up language designed to store and transport data. It is not designed to display data. XML came into existence in late 90's. it was created to provide an easy way to use and store self describing data. XML is not a replacement for HTML. It is platform independent and language independent.

The main benefit of xml is that you can use it to take data from a program like Microsoft SQL, convert it into XML then share that XML with other programs and platforms. You can communicate between two platforms which are generally very difficult. Many corporation use XML interfaces for databases, programming, office application mobile phones and more. It is due to its platform independent feature.

With XML, your data can be available to all kinds of “reading machines” (Handheld computers, mobile phones, smart devices, voice machines, news feeds, etc), and make it more available for blind people, or people with other disabilities.

XHTML (eXtensible HyperText Markup Language) is a markup language which is written in XML. It is a hybrid technology between HTML and XML that combines the functionalities of both to become powerful and efficient. HTML is used for the presentation of the data, while XML is used for carrying the data. It was developed by World Wide Web Consortium (W3C), which is an international organization that sets standards for the World Wide Web (WWW). It was designed to help web developers to make the transition from HTML to XML. It is specifically designed for net device displays.

Activity B

A Principal wants to make a web site for the information of parents so that they can get information of their progress of their ward in the school as well about various schemes of the school.

1. Make a diagram for the website
2. Which language would you prefer for that website

.....

.....

.....

.....

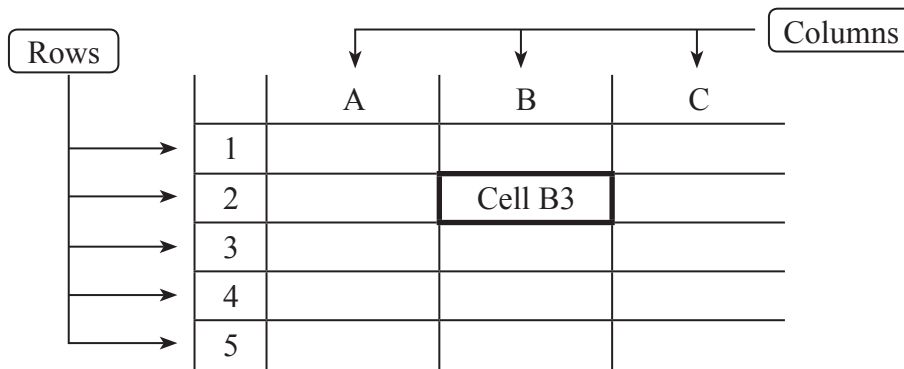
.....

11.8 OVERVIEW AND FEATURES OF EXCEL

Microsoft Excel is a program specially designed to enter organized data as well as to analyze and present the data attractively. Excel is one of the most versatile and popular spreadsheet programs. It serves as an, electronic pad for accountants. It an easily perform simple as well as complex mathematical operations. Spreadsheet is a simple worksheet consisting of rows and column in which any data can be entered e.g. report cards of students are manual spreadsheet. An electronic spreadsheet instead of being on papers, it is on the computer screen.

Spreadsheets are used for performing calculations, recalculating results if any data stored in them is changed, creating financial reports, comparing reports etc. A very useful feature of spreadsheet is its ability to create groups. It helps you establish relation-ship between two or more sets of data and easily understand the trends of data changes.

A spreadsheet consists of rows and columns, which combine to form cells. A Cell is a box where we can enter data. Column form the vertical lines of calls while rows form the horizontal line of cells. Cell is an intersection of rows and columns. To describe the location or address of a cell, we have to write the names of the column and the row whose intersection has created this cell.



Cell B3 is location of address of this cell, which means column B, & row 3

Labels are the headings, which we enter in cell. Values are the number on which calculations are performed. Formulas and functions are elements that perform the desired calculations on the values.

	A	B	C	D	E	Value
1		Profit in months				
2						
3	Export	1,200	1,600	1,479		
4		2,200	1,785	2,100		
5						
6	Total	3000	C3+C4	3,579		

Label formula or function

Formulas, which are present within the spreadsheet itself, are called functions. We can also put up our own formulae in any cell. In the above figure cell B6; C6 and D6 contain the formula, which total the values of export and import.

A spreadsheet can be designed in a number of formats. Basically, designing is concerned with how a number should look on the screen. The various designing formats are comma, percent, currency, scientific. The data in a cell may be right aligned, left aligned or centered.

Exam	Left-aligned
Exam	Right-aligned
Exam	Centered

Spreadsheets allow us to use different fonts and styles to emphasize headings.

Graph is yet another powerful feature of spreadsheet. The data stored in the spreadsheets can be converted into graph so as to study the relationship various data. The spreadsheet program has the facility to draw various graphs viz. Bar, Pie, 2D and 3D graphs.

Spreadsheet allows the user to protect their workbooks by using a password from unauthorized access to their information.

11.9 SUMMARY

In this unit we have explained the concept and fundamental of programming languages. We have also explained the different type of data types and looping statements available in Visual Basic (VB.Net) and Java programming language. We also learn how to compile and run programs in above programming languages.

11.10 UNIT END EXERCISES

1. Explain the concept of programming language. Explain identifiers, constants, expressions, and library functions.
2. Give points to support that Visual Basic is an excellent development tool. Highlight important features of VB.
3. Java is not 100% pure object oriented language? Do you agree or not? Justify your answer.
4. What do you understand by HTML? Present the features of HTML.
5. “Excel is a versatile spreadsheet package. It can do wonder for accountants”. Comment.
6. “With so many ready made and customized software available-The need for a manager is to learn to use them effectively rather than learn to program them.”. Do you agree?
7. What makes Java an almost perfect web programming language? What are its disadvantages?
8. Differentiate between HTML, XML, and XHTML.

11.11 REFERENCES & SUGGESTED FURTHER READING

Sebesta, *Concepts Of Programming Languages, 4e*, Pearson Education

Terrence W. Pratt, Marvin V. Zelkowitz , *Programming Languages: Design And Implementation, 4th Ed.* , Prentice-Hall Of India

Tucker, Allen, *Programming Languages Principles and Paradigms*, Tata Mc Graw-Hill

<https://dotnet.microsoft.com/learn/dotnet/what-is-dotnet>, October 10, 2021

https://www.tutorialspoint.com/vb.net/vb.net_overview.htm, October 12, 2021

[https://techvidvan.com/...](https://techvidvan.com/) October 16, 2021

<https://docs.oracle.com/javase/7/docs/technotes/tools/index.html> October 18, 2021

<https://www.javatpoint.com/> <https://www.javatpoint.com> October 19, 2021

<https://html.com/> October 19, 2021

Block

4

SUPPORT SYSTEMS FOR MANAGEMENT DECISIONS

UNIT 12

Database Resource Management	217
-------------------------------------	------------

UNIT 13

Data Warehousing and Data Mining	234
---	------------

UNIT 14

Artificial Intelligence and Decision Support Systems	270
---	------------

UNIT 15

Emerging Trends in IT	291
------------------------------	------------

BLOCK 4: SUPPORT SYSTEMS FOR MANAGEMENT DECISIONS

Unit 12: Database Resource Management begins with defining data as a data as organizational resource. It describes the concept of a Database Management System (DBMS). It spells out how data is important to an organization. Role of a database administrator in DBMS is also explained. It closes with identifying the factors that lie behind the success of Data Ware Housing.

Unit 13: Data Warehousing & Data Mining highlights the importance of building an Enterprise Data Warehouse. It tells about what a data warehouse is and the types of data it contains. It makes you understand the design decisions involved in data warehouse design and the architecture of a data warehouse and its various components. You will be able to understand the importance of external data feed from the World Wide Web to the data warehouse. You will also be familiar with multi-dimensional data analysis and on-line analytical processing. Finally it enables you to understand how to apply data mining tools for discovery of hidden relationships and patterns from business data

Unit 14: Artificial Intelligence and Decision support systems explains what artificial intelligence, expert systems and knowledge base systems mean. It portrays Neural Networks, Genetic Algorithms, and Fuzzy Logic. You will be able to appreciate the efforts made in creation of these systems. It also identifies and discusses their advantages in business applications.

Unit 15: Describe the competitiveness of information and communication technology (ICT) with respect to modern business practices and their impact on economies;

UNIT 12: DATABASE RESOURCE MANAGEMENT

Structure

- 12.1 Introduction
- 12.2 Objectives
- 12.3 Data as Organizational Resource
- 12.4 Organizing data
- 12.5 Database Management System and its Components
- 12.6 Data Modelling
- 12.7 Attributes in Tables
- 12.8 Developments in Database Technology
- 12.9 Database Administrator and the Role
- 12.10 Database and Data Warehousing
- 12.11 Summary
- 12.12 Unit End Exercises
- 12.13 References & Suggested Further Reading

12.1 INTRODUCTION

Every organisation today uses respective database for their business. An educational institution uses a database to maintain student records, libraries use a database system to keep track of circulation of reading materials and to provide various types of indexing to material by subject/author/title etc and a bank uses database system to keep records of its customers, employees, daily transactions to & from various accounts, loans etc. and many more. Similarly it happens in production houses, warehouses, health, R&D etc for respective businesses. Integrated application software and World Wide Web (WWW) with internet have made it more adaptive and simpler in maintaining databases. Businesses houses normally use a database to keep records of its employees, sales, production etc. also connecting their branches across globe. All those who keep a collection of records for a common purpose, uses a database. In order to build and use a data base, a software management system which is known as Data Base Management System (DBMS) is used for developing the corresponding application software for the purpose. ERP, CRM, SCM etc are the integrated applications readily available which need only customisation to suit the business requirements of the organisation willing to use them. In such case, application software is not required to be developed. As such, DBMS is a collection of software programs, which allows developing application software for creation, storage, modification, and manipulation of data from a database. Application software is a layer of software between the physical database stored in computers and the user. DBMS is layer between application software and the operating system running on computers.

12.2 OBJECTIVES

After reading this unit, you should be able to:

- Describe the concept of a Data Base Management System (DBMS);
- Describe the importance of data in organizations;
- Explain the role of Data Base Administrator (DBA); and
- Identify the factors responsible for the success of Data Ware Housing.

12.3 DATA AS ORGANIZATIONAL RESOURCE

Before the computer came into existence, there were many limitations associated with the physical handling of documents and human processing. Computers came into existence to speed up the data processing. Computers helped to manage data.

We are living in the age of Information Technology. The ability to acquire accurate and timely data, managing data efficiently, is all through which an organization succeeds. The development in hardware has catalysed rapid development in software and now we are having more sophisticated software to handle and maintain data.

The term data and information are often taken as to mean the same thing. Data are the details about factories, outlets, staff, competitors, customers and suppliers. Data is also kept for monitoring the activities and processes in business. Once businesses collect these details, information comes into picture. Information is the collection of meaningful facts that are derived from the data. Information is significant and relevant to the user unlike data, which has no meaning alone. Information leads to decision and thus an appropriate action. In fact, information consists of only the data that is useful, meaningful and needed. You come across many information systems in your daily life like banking information system, online shopping, ticketing and reservations systems etc. Now what you get out of these information systems is pure information and these systems work out this information on the basis of data. In other words, data is raw and information is refined. The exact form of the refinement depends on the type of application one is dealing with.

The data must be accurate, timely and relevant. It is desired that any information system should have accurate data to work on. Inaccurate data howsoever well analysed will hardly be useful for any decision-making. Although collection of accurate data is costly and time consuming, every effort should be made to make it as accurate as possible. One may strike a balance between the cost of processing and value of accuracy. The data should also be timely. If the right data is not available at the right time then it is worth little use. For example if the updating of electoral lists cannot take place before the schedule of elections then it is of no use to the elections. The timeliness of the data is detrimental of success of any information system. The data should be relevant. For better understanding of the organisations and their information needs data has to be high on the relevance factor. There are many decision-support systems (DSS) and executive support systems that are present in today's scene, but the general feeling is that

often the data is generating accurate and timely information, that is not very relevant. However, with emergence of high-end computing machines, sophisticated data capturing devices and other technology driven tools it is possible to make data available in larger quantities than ever before at lower storage cost, increased processing speeds, higher capacity communications and increased variety of information formats.

12.4 ORGANIZING DATA

In earlier times, data processing was done manually. Organizations used to appoint a large number of people called clerks. The information technology devices used at that time were forms, ledger books and basic mechanical adding machines. The results of such manual operations were obtained at a time when the information was almost out of date (e.g. census). Then some systems were invented in which processing was mostly mechanized e.g. Hollerith Tabulation System. In such systems data was recorded in the binary form of holes in cards, using a cardpunch. These stacks of cards could be sorted and tabulated. IBM and Remington Rand led the development of punched card technology.

Things have changed considerably with the advent of computers. There are a few terms, put up below, that you need to know about the data organization, although they are not new these days in this era of IT. As they are the basics in data organizations hence they are put here for maintaining the continuity and better understanding.

Bit: A bit (short for binary digit) is the smallest unit of data in a computer. A bit has a single binary value, either 0 or 1. Although computers usually provide instructions that can test and manipulate bits, they generally are designed to store data and execute instructions in bit multiples called bytes. In most computer systems, there are eight bits in a byte.

Byte: In most computer systems, a byte is a unit of data that is eight binary digits long. A byte is the unit most computers use to represent a character such as a letter, number, or typographic symbol (for example, “g”, “5”, or “?”). A byte can also hold a string of bits that need to be used in some larger unit for application purposes

Field: A field is an area in a fixed or known location in a unit of data such as a record, message header, or computer instruction that has a purpose and usually a fixed size. In some contexts, a field can be subdivided into smaller fields. In a database table, a field is a data structure for a single piece of data. Fields are organized into records, which contain all the information within the table relevant to a specific entity. For example, in a table called customer contact information, telephone number would likely be a field in a row that would also contain other fields such as street address and city. The records make up the table rows and the fields make up the columns. Fields are also called as elements or data items.

Record: In a database, a record (sometimes called a row in a table) is a group of fields within a table that are relevant to a specific entity. For example, in a table called Customer-contact, a row would likely contain fields such as: Customer-ID number, Customer-name, Street address, City, Telephone number, Mobile number, eMail address and so on.

File: In data processing, using an office metaphor, a file is a related collection of records. For example, you might put the records you have on each of your customers in a file. In turn, each record would consist of fields for individual data items, such as customer name, customer number, customer address, and so forth.

Database: A database is a collection of data that is organized so that it can easily be accessed, managed, and updated. In one view, databases can be classified according to types of content: bibliographic, full-text, numeric, images, audio, video as so on.

In relation to database, an entity means a person, place, or thing that we wish to collect information on (e.g. customers). The word root is from the Latin, *ens*, or being, and makes a distinction between a thing's existence and its qualities. Attribute is a characteristic of an entity (e.g. customer's salary, customer's address). In the application software developed using a Data Base Management System (DBMS), an attribute may describe a component of the database, such as a table or a field, or may be used itself as another term for a field. Key Field is a special field that uniquely identifies a single record (e.g. customer's registration number). It can be a collection of fields.

In a DBMS, files are either organised sequentially (one record after another, used for batch processing) or organised in an indexed sequential form (in sequence, but records can be directly accessed using an index) or in the form of direct or random form (records located using a key field generated by a mathematical formula).

12.5 DBMS AND ITS COMPONENTS

We have discussed earlier that an organization must have accurate and reliable data for effective decision-making. For this, the organization maintains records by combining the data from different sources in the organization. DBMS is the tool required for managing the data in the database. In other words, DBMS is a complex piece of software that facilitates a flexible management of the data. Through DBMS we can create, access, monitor, store and modify the database. Through DBMS data can be made available to all users and redundant (duplicate) data can be minimized or completely eliminated. DBMS also makes possible for an organization to prevent important data access from unauthorized users by providing the security to the database at different levels. Some of the DBMS that are used are Oracle, SQL Server, MySQL, Microsoft Access.

The DBMS allows users to access data from the database having no knowledge of how data is actually stored in it. The process is much the same as ordering a menu in the restaurant. A customer simply orders for the food to a waiter and waiter serves the specified order. A customer only checks the menu for the desired items and need not know how the items are arranged in a restaurant. Similarly, the database users need not know how the data is stored instead they need to know only what they require and the DBMS takes care of retrieving the required data on its own.

Components of DBMS

Computer is just a machine. It does not do anything on its own. For each and everything some or the other components are required in terms of

software to communicate and get things done by computer. DBMS is a software having components for various uses in order to build and operate databases. Such components are as below:

DML Pre-compiler (Data Manipulation Language Pre-compiler): As its name specifies, this is a compiler that converts DML statements (Statements that allows the users to manipulate the database) in an application program into normal procedure calls in the host language. This pre-compiler must interact with the query processor (discussed later) in order to generate the appropriate code. It is also known as DML Processor or Pre-processor.

DDL Compiler (Data Definition Language Compiler): The DDL compiler converts the data definition statements (statements that define the tables, database etc.) into a set of tables. These tables contain information in the form that can be used by other components of the DBMS.

File Manager: File manager manages the structure and space of the file on disk storage. This also locates the block in which the required record exists and requests from the disk manager for this block containing the required record and finally provides the required record to the data manager.

Disk Manager: As specified earlier, disk manager provides the block or page that the file manager asks for. The disk manager is a part of Operating System. Disk manager does all the physical input and output operations. Interface program of DBMS establishes the link between DBMS and Operating System.

Data Manager/ Database Manager: It is also called database control system. It is a software component of the DBMS or we can say a program module that acts as an interface between the data stored in the database and the queries submitted to the system. Data Manager converts the queries from the user into the file system. It is actually responsible for the storage, retrieval and manipulation of the data in the database. Besides these responsibilities data managers also takes care of maintaining integrity. This means that the appropriate data should be stored in the database. For example, date of joining of an employee should not be less than the current date or the age (date of birth) of an employee for a particular post. If these constraints are specified in the database, then database manager takes care of them on its own. Database Manager also checks for an unauthorized user accessing the database. Database Manager does not allow any user to access the database without having the permission granted.

Query Processor: The Query Processor takes care of the queries by the database user. The database user when uses the Data Manipulation Language (DML) for retrieving the data, the Query Processor converts it into that form that could be sent to the Data Manager, so that Data Manager can execute it.

Data Dictionary: A data dictionary contains the information/ data about the data. In other words, it contains the met data. A data dictionary includes all the database descriptions, entity/ table descriptions, attributes etc. Apart from these, data dictionary includes the information about the users also, which tables are used by which programs, authorization of users etc. A data directory/ dictionary is just like a database and its cost depends on how complex it is? Complexity increases as the more and more facilities are

added. Thus, more complexity leads to increase in cost. Just like an index helps us to find the chapter easily, data dictionary helps DBA for designing, implementing and maintaining the database and searching for the desired record in the database. It also helps the managers and end users in their project planning.

12.6 DATA MODELLING

We have discussed earlier that an organization needs to store data to process and have the desired information, and this is done through database. Data models refer to different views of data and the underlying relationships among them. Data modelling is the process of creating a data model for an information system. It gives a visual representation of either a whole information system or parts of it. It shows connections between data points. It illustrates the types of data used and stored within the system, the relationships among these data types, the ways the data can be grouped and organized and its formats and attributes. This process is used to define and analyze data requirements needed to support the business processes within the scope of corresponding information systems in organizations.

Data models are built around business needs. Rules and requirements are defined upfront through feedback from business stakeholders so that they can be incorporated into the design of a new system or adapted in the existing one. The process begins by collecting information about business requirements from stakeholders and end users. Business rules are then translated into data structures to formulate a concrete database design. A data model can be compared to a roadmap, an architect's blueprint or any formal diagram that facilitates a deeper understanding of what is being designed.

Data modelling techniques and methodologies are used to model data in a standard, consistent, predictable manner in order to manage it as a resource. Data modelling makes it easier for developers, data architects, business analysts, and other stakeholders like program testers, manual writers, IT package selectors, engineers, managers, related organizations and clients to view and understand relationships among the data in a database or data warehouse. Further benefits are that it can:

- reduce errors in software and database development
- increase consistency in documentation and system design across the enterprise
- improve application and database performance
- ease data mapping throughout the organization
- improve communication between developers and business intelligence teams
- ease and speed the process of database design at the conceptual, logical and physical levels
- help manage data as a resource
- be used to integrate information systems
- be used in design of data warehouses and data repositories.

Data models can generally be divided into three categories. The process starts with a conceptual model, progress to a logical model and concludes with a physical model. The three data models are as below:

Conceptual data models: They provide view of what the system will contain how it will be organized, and which business rules are involved. Conceptual models are usually created as part of the process of gathering initial project requirements. Typically, they include entity classes (defining the types of things that are important for the business to represent in the data model), their characteristics and constraints, the relationships between them and relevant security and data integrity requirements.

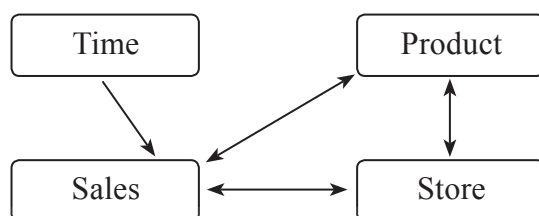


Figure 12.1 Conceptual Data Modelling

Logical data models: They provide greater detail about the concepts and relationships in the domain under consideration. They indicate data attributes, such as data types and their corresponding lengths, and show the relationships among entities. Logical data models don't specify any technical system requirements like storage requirements etc.

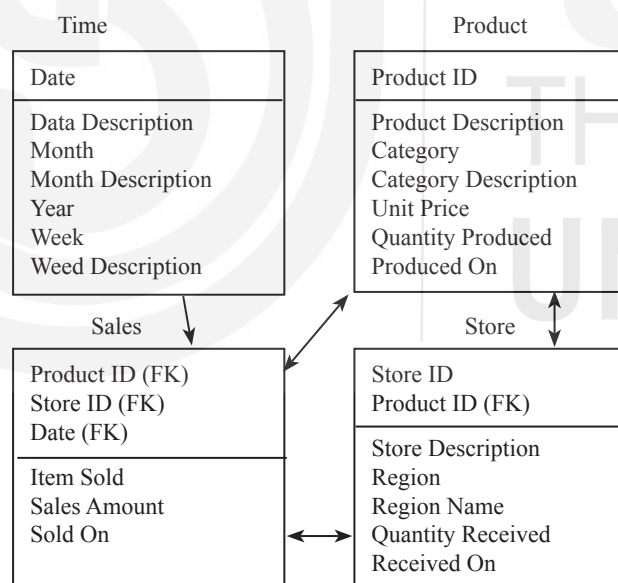


Figure 12.2: Logical Data Modelling

Physical data models: They provide a schema for how the data will be physically stored within a database. As such, they're the least abstract of all. They offer a finalized design that can be implemented as a relational database, including associative tables that illustrate the relationships among entities as well as the primary keys and foreign keys that will be used to maintain those relationships. Physical data models can include database management system (DBMS)-specific properties, including performance tuning.

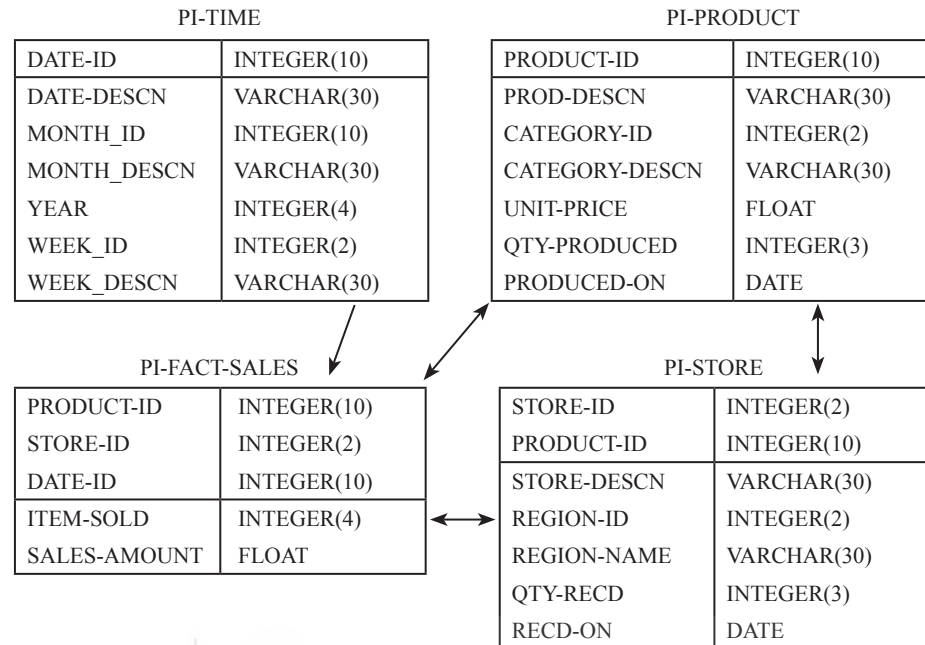


Figure 12.3: Physical Data Modelling

Numerous commercial and open source computer-aided software engineering (CASE) solutions are widely used today, including multiple data modeling, diagramming and visualization tools. Some examples are - **erwin Data Modeler, Enterprise Architect, ER/Studio**, and free (open source) data modeling tools such as **Open ModelSphere**.

All the above data models can be broadly classified into three:

- Record based data models
- Object Oriented data models
- Dimensional data models.

The record based data modelling has been in practice, since beginning of application software development using database approach. The three sub-models are – Hierarchical, Network, and Relational models as briefed below:

Hierarchical models: - These are the early data models used in 1970's. Hierarchical models capture the intuitive hierarchy of the data elements. The early generation of large DBMS e.g. IMS (Information Management system) by IBM belongs to the hierarchical data models and were widely used especially in banking. Though this approach is less efficient than more recently developed database models, it's still used in Extensible Markup Language (XML) systems and geographic information systems (GISs).

Network models: - Since hierarchical models are unable to represent data items that existing at two different level of hierarchy, network models were proposed. The notable systems built using this model were ADABAS, IDMS (Integrated Database Management System), and DBMS-10. They were found to be more powerful and were used in various sectors like telecom, production houses etc.

Relational models: - Though network models were quite powerful, they lacked in elegance and were not simple to use. DBMS with relational model is also known as RDBMS. The systems built on this data model were dBase, Xbase and ORACLE. Almost all commercial systems presently available like Oracle 11G, Oracle12C, Oracle 18C, and the latest as of now (in 2021) Oracle 19C, SQL Server, MySQL are built on the relational models. Data modelling in RDBMS is achieved in two levels: -

1. ER modelling that builds the Entities and provides Entity Relationship diagrams
2. Normalization, which removes the redundancies.

Object-oriented data models came into practice because of object-oriented programming and it became popular in the mid-1990s. The “objects” involved are abstractions of real-world entities. Objects are grouped in class hierarchies, and have associated features. Object-oriented databases can incorporate tables, but can also support more complex data relationships. This approach is mostly used in multimedia and hypertext databases. The approach is however not as simple as the Relational data modelling in which case ER diagram very simply depicts all the entities, attributes, and relationships between entities. Object Oriented Data (OOD) modelling has advantage of getting more redundancies removed and hence uses less storage space as compared to ER model.

Dimensional data models are designed to optimize data retrieval speeds for analytic purposes in a data warehouse. While OOD and ERD data models emphasize efficient storage, dimensional models increase redundancy in order to make it easier to locate information for reporting and retrieval. This modeling is typically used across OLAP systems. OLAP (for *online analytical processing*) is software for performing multidimensional analysis at high speeds on large volumes of data from a data warehouse, data mart, or some other unified, centralized data store.

The ER data model is the most common data model which allows us to represent the relationships among the entities (which are called as objects in OOD model). It uses following three concepts to represent itself graphically. Although we have discussed them earlier, we revisit them here once again:

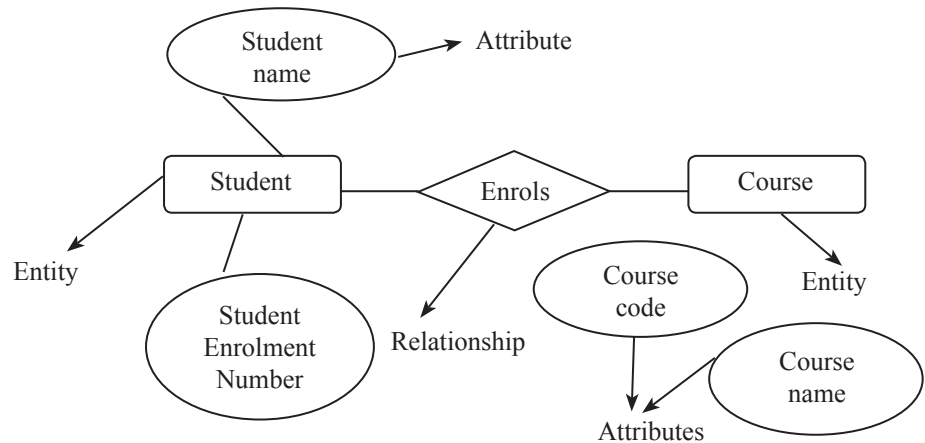
Entity: Any real-world object that has certain properties (attributes of its own) and this object are uniquely identified by the system on the basis of these. This is represented by a rectangle with the entity name specified in the centre.

For e.g.: - student is an entity.

Attribute: - Attribute is the properties of an entity, like a student entity can have attributes such as student-name, student-roll-number, student-class etc. Attributes are placed inside the circles and attached to the entities and relationships.

Relationship: - Relationship specifies the meaningful relation between two entities. This relationship can also have attributes. A rhombus represents

these with a relation specified in it. For e.g.: - if we have student and course as entities, then we can relate these entities as student enrolls for a course.



Degree of Relationship (DOR): - This specifies the occurrence of entity with other entity.

One to One (1:1): - In this, there can be almost one related occurrence for each entity. For ex: - A single manager manages one department and one department can be managed by a single manager. It is represented as follows.



One to Many (1:N): - Here, for one occurrence of the first entity, there may exist many occurrences of the second entity and for every occurrence of the second entity, there exists only one occurrence of the first entity e.g.



Here one manager supervises many employees and every employee reports to only one manager.

Many to Many (M: N): - In this degree of relationship, for one occurrence of the first entity, there may exist many occurrences of the second entity and also for every occurrence of second entity, there exists many occurrences of the first entity e.g.



One employee can work for many projects and many employees can handle one project.

Normalization is another concept in data modelling. Normalization is a process of converting complex data into simpler form without any loss of information. The Normalization technique ensures that there is no dependent and duplicate data when the E-R model has been made. Normalization first

of all, converts the data into tables or relations. These tables are checked for redundancy. Finally the form is converted to a database definition. Why do we need normalization? Normalization improves the database design. Whenever the design is modified, we need to re-organize the data. Normalization reduces the need of reorganizing the data. Process of normalization removes all undesirable consequences that may occur due to inserting, updating and deleting values from the tables. Normalization reduces the unnecessary repetition of data (redundancy) that causes the problem with storage and retrieval of data. Poorly designed databases have several data management difficulties e.g.

Consider a relation

LIBRARY (MNAME, MADDR, B_TITLE, ISS_DATE, DUE_DATE)

This relation has the following problems: -

1. Here duplicate data exists because member address will be repeated if a member issues more than one book.
2. Insertion problems/ anomalies: - If a member does not issues a book then we cannot record the address of a member. Here MNAME & B_TITLE are both necessary and hence we cannot leave any one of them as blank.
3. Updating problem: - If a member has issued two or more books, then address is necessary in all and if a member's address needs to be updated, then all the records needs to be updated.
4. Delete anomalies: - Here if we delete the issue details, then address would not be there.

But all the above problems can be eliminated by normalization. If this single relation is broken down in two relations:

MEMBER_DETAILS (MNAME, MADDR)

MEMBER_TRANS (MNAME, B_TITLE, ISS_DATE, DUE_DATE)

But this requires a join between these two relations, which is very expensive.

Now lets look into several normal forms with the help of an example: -

Table 12.1: Un-normalized data

Rollno	Name	Subject	Marks
210023456	Amit	English	65
		Hindi	72
		Maths	70
210023469	Seema	English	70
210023477	Anjali	English	54
		Hindi	60

To make this un-normalized data into 1 NF (First Normal Form), we have to make each cell containing one value. And all the repeated information should be removed.

Table 12.2: First Normal Form

Rollno	Name	Subject	Marks
210023456	Amit	English	65
210023456	Amit	Hindi	72
210023456	Amit	Maths	70
210023469	Seema	English	70
210023477	Anjali	English	54
210023477	Anjali	Hindi	60

The data is still redundant in this form. Here combination of two keys (composite key), Rollno and Subject, is a primary key (PK, that uniquely identifies a record). But the attributes of this table depend on the part of the PK. Rollno and subject determines marks, Rollno determines Name, Name has no dependency on the attribute subject. This may lead to following problems:

Insertion: the name of a subject cannot be recorded until a student gives any exam.

Updation: Rollno and name is repeated several times. If name of the student is to be changed, then at every place the name needs to be updated otherwise may lead to inconsistencies.

Deletion: if a student has not given any of the three exams, then the name of that particular student will be deleted.

So, for overcoming these problems, this first normal form needs to be decomposed and converted into Second Normal Form (2NF) without any loss of information.

Table 12.3: Relation/Table in 2NF

Roll no.	Name	Roll no.	Subject	Marks
210023456	Amit	210023456	English	65
210023469	Seema	210023456	Hindi	72
210023477	Anjali	210023456	Maths	70
		210023469	English	70
		210023477	English	54
		210023477	Hindi	60

A relation is said to be in 2NF, only if it is in 1NF and every attribute is dependent on the whole key and not just a part of it.

Other major normal forms are 3NF, BCNF & 4NF. Third normal form or 3NF is used to prevent loss of information & dependencies preserving decomposition. BCNF is the advance version of 3NF. It is stricter than 3NF. Fourth normal form or 4NF is used to preserve multi-valued dependency, which is essentially a constraint. There are several other normal forms like project-join NF, domain-key NF. We are not describing them as it is not within the scope of this course.

Activity A

Prepare a comparative chart of the strengths and weaknesses of two DBMS products in the market e.g. MS-ACCESS and Oracle.

.....

.....

.....

.....

12.7 ATTRIBUTES IN TABLES

In ER model, each Entity and Relation has a corresponding Table. For storing the data in a database system, each column in the table have a certain type of data. For example when we create a table for employee's personal data, then we may have different fields (attributes) like name, address, phone, age etc. In this if we take the attribute "name" then this would be of character type and "age" & "phone" are of numeric type.

Sometimes an attribute of a table T takes the values from a set D that is known as a domain. A domain is defined in the cases when we want an attribute to take some exact set of constants. For instance, the domain of the attribute 'age' for a particular post that a candidate can apply for, might be defined as consisting of all the integer numbers between 18 and 25. On the other hand, the domain of the marks column, might be defined as all the numbers between 0 and 100, with at most one non-zero digit after the decimal point (assuming the maximum marks to be 100). The significance of declaring a column of a table to have a particular type (or domain) is the ability to compare the values of two different column rests on this declaration. A particular column in a table must contain similar data, which is of a particular type. Besides storing character, numeric, we can also store data such as graphics, sound etc. Whenever we have the bio data of employees or students or candidates, generally we store their photographs. In other applications like for passport or in Aadhar/PAN/ATM card etc. biometrics is essentially used. In such cases, we can use graphics as a data type for photograph and biometrics columns.

12.8 DEVELOPMENTS IN DATABASE TECHNOLOGY

DBMS consists of a collection of interrelated data that is called a database and a collection of programs to manage & access the data. An RDBMS is a collection of tables, each of which is assigned a unique name. Each table consists of a set of fields and stores a large set of records. A data warehouse is a repository of data collected from multiple sources, stored under a unified schema that resides at a single site. There have been a lot of developments in the field of database technology. These developments include handling of spatial data such as maps; engineering design data such as design of buildings, components and circuits; multimedia data such as text, image, video and audio data; time related data such as historical records or stock

exchange data and the World Wide Web data. Han & Kamber (2001) had described each of these advanced databases:

1. *Object Oriented Databases*: In object oriented databases each entity is considered as an object. These objects can be customers, suppliers, employees or items. Objects that share a common set of properties can be grouped into a Class that represents properties common to the Class.
2. *Object Relational Databases*: In object relational data model much is the same as the basic relational data model but for the fact that object relational databases also handles complex data types and complex object structures.
3. *Spatial Databases*: These databases contain geographic databases, clip design databases, medical and satellite image databases. These data are represented in raster format, consisting of n-dimensional bit maps or pixel maps as well as vector format, consisting of basic geometric shapes such as points, lines and polygon etc.
4. *Time Series & Temporal Database*: These databases contain time-related data. A temporal database usually stores relational data that include time-related attributes. A time series database stores sequences of values that change with time.
5. *Text & Multimedia Databases*: Text databases contain word description for objects (these could be long sentences or paragraphs). Multimedia databases store images, audio and video data. These databases are used in applications such as picture content-based retrieval, voice mail systems, video-on-demand systems & World Wide Web etc.
6. *Heterogeneous Databases*: Objects in one component database may differ greatly from objects in other component databases, making it difficult to assimilate their semantics into the overall heterogeneous database. There are special methods for analysis of such databases.
7. *World Wide Web Databases*: World Wide Web has data objects linked together to facilitate interactive access. Users seeking information of interest traverse from one object, via links, to another. These databases are also called as virtual databases. It is mostly used by organisations for gathering secondary data for business analysis.

Activity B

Now that you know about various types of data, you can encounter in day-to-day life. Take the case of an organisation of your choice and try to identify two examples of each of these data types. Also, comment on the usefulness of analysing these data types for the organisation

.....

.....

.....

.....

12.9 DATABASE ADMINISTRATOR AND THE ROLE

Large and complex databases need a lot of maintenance to work in a proper manner. A database administrator (DBA) does this maintenance. DBA is the supervisor of the database and hence have the highest level of expertise. DBA centrally controls the database and administers the overall user community. DBA also takes care of the data and controls the database structures. DBA is responsible for accessing the database so that an unauthorized person should not access the database. DBA stores the permissions granted to the users created to verify that a particular user can only perform a given operation on the database. DBA can also revoke the permission if he finds that the user or user group needs no more access to that particular portion of the database.

DBA account is sometimes known as the system account that has the powerful capabilities that other regular database users don't have. Sometimes when more than one user uses the same resource such as the data, there may exist the conflict but the DBA seeks for the compromise and takes care of the conflict for the benefit of the organization. Hence, DBA also has the responsibility for maintaining the integrity of data. This means that only accurate data can be inserted into the database.

Duties of DBA include the existence control, definition control, quality control, update control, access control, storage control etc. The DBA is also responsible for recovering the database from failures that may be due to human disasters, software and hardware causes or natural causes with the minimal loss of data. DBA changes the database according to the changes or growth in the organization. Hence, DBA is said to be a super user who is responsible for monitoring and improving the performance and operations to maintain the database activities. The database dictionary is one of the most important tools for the database administrator. It is used to maintain information related to various resources. It contains metadata i.e. data about data of all applications running on single or multiple servers.

12.10 DATABASE AND DATA WAREHOUSING

Database is collection of data on a subject. The word Data Warehouse was first introduced by Bill Inmon, in 1990. According to him, "A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process'. You will read in detail about data warehousing in the next unit. A data warehouse is a central repository for significant pieces of the data that are collected by various business systems. Data from various online transaction processing applications and other sources is organized on the data warehouse for use decision-making process. A data warehouse is used in applications like data mining, Web mining and a decision support system (DSS). Data Warehouse enables the executive, manager or the analyst to make better and faster decisions. It presents the right information in the right place at the right time with the right cost in order to support the right decision. According to

Inmon, a data warehouse should firstly, have a common format meaning that data that goes into a warehouse should have a common data format. Since a data warehouse is made out of a combination of various organizational databases of the company, every care should be taken to maintain uniformity. Secondly, since different company functions such as Sales, Marketing, HR, Finance, and Production have some variables common in a database a data warehouse should be able to use these data pertaining to different functions. Thirdly, since data often contain time as a dimension and is updated into a data warehouse periodically, a data warehouse should be efficient in computing all time dependent aggregations again and again. Finally since data warehouses are never deleted or updated they should be able to sustain large chunks of data (may be in terabytes or beyond). A data warehouse grows all the time as new data keeps on adding. In practice, a data warehouse is often “resynchronized” to keep its volume manageable. Resynchronization means that data is kept in periods and as and when a new period is added, the oldest one is deleted.

These are some very preliminary concepts about data warehousing. You will read about it in more detail in the next unit.

12.11 SUMMARY

Data consists of facts and details about things, activities, transactions etc. Data can be stored as a resource that can be drawn later on to produce information for the people and activities that need it. In this unit we have discussed about the prominent features of a modern database management system. Understanding a DBMS is very significant for a manager dealing with information systems. You have learnt about modelling data and different data types. You saw the developments in the field of database technology. This shows that almost any kind of data can be handled, be it a music database or a database of photographs stored on the computer. The unit closes with giving you a glimpse of success factors in data warehousing, which you would read in detail in the next unit.

We have limited our scope in the sense that we would not like to make you a programming expert but would be more interested in making you to capable of applying these concepts for deriving business advantage.

12.12 UNIT END EXERCISES

1. What do you understand by “data” and “information”? Give three examples each to distinguish between data and information.
2. Explain the meaning of the terms database, database management system and database administrator in your own words.
3. Explain what do you understand by database management systems? Describe its components.
4. What are data models? How you would classify the data models? Differentiate between various data models with the help of an example.
5. What is normalization and why it is used? Explain with the help of an example how would you derive normal forms.

6. Discuss different types of data in database system. What are the developments in database technology with respect to data types?

12.13 REFERENCES & SUGGESTED FURTHER READINGS

Combs, Martin R, *Information Systems for Business Management*, Pitman Publishing, 128, Long Acre, London

Han Jiawei & Kamber Micheline, *Data Mining: Concept & Techniques*, Morgan Kaufmann, Elsevier Science, India

Hicks Jr, James O., *Information Systems in Business: An Introduction, second edition*, West Publishing Company, St. Paul, MN

[http://whatis.techtarget.com& http://searchdatabase.techtarget.com/](http://whatis.techtarget.com&http://searchdatabase.techtarget.com/) for definitions of terms.

Site accessed on 19/05/2004.

Kroth Henry F, Silberschatz A., *Database System Concepts*, McGraw Hill International Edition.

Sadgopan S., *Management Information Systems*, Prentice-Hall India Private Limited, New Delhi, India

<https://www.ibm.com/cloud/learn/data-modeling> Site accessed on October 20, 2021

UNIT 13: DATA WAREHOUSING AND DATA MINING

Structure

- 13.1 Introduction
- 13.2 Objectives
- 13.3 Operational and Informational Databases
- 13.4 The Data Warehouse
- 13.5 Data Warehouse Schema
- 13.6 Metadata
- 13.7 Data Warehouse and the Web
- 13.8 On-Line Analytical Processing (OLAP)
- 13.9 Data Visualization
- 13.10 Data Mining
- 13.11 Summary
- 13.12 Unit end Exercises
- 13.13 References and Suggested Further Readings

13.1 INTRODUCTION

Use of IT in all the fields and business sectors is well known today. Developments taking place have made information communication very easy. A smart phone is sufficient today to generate and disseminate information across without barriers. Same information is possible today to make available over many channels through eMails, social media, messaging systems and so on in unified way. Automation of day-to-day activities, availability of internet in every nook and corner, ease of governmental policies is starting businesses, use of IT by one and all in business are producing more and more volume of data which are used today for various analytical purposes. Such analyses are also used for proactive strategic decisions. Availability of resources like cloud computing SAAS (Software As A Service), PAAS (Platform As A Service) etc have made it possible to dig out data from around the globe. In this Unit, we will discuss the concept and use of *data warehouses* in terms of how they have evolved, matured, and taken the world of business by storm.

13.2 OBJECTIVES

After reading this unit you should be able to:

- Identify the types of data in data warehouse;
- Describe the design decisions involved in data warehouse design;
- Explain the importance of building an Enterprise Data Warehouse;
- Depict the architecture of a data warehouse and its various components;
- Describe various aspects of Star and Snowflake schemas;

- Define and explain the concept of metadata in the data warehouse;
- Enumerate the importance of external data feed from the World Wide Web to the data warehouse;
- Describe the process of data analysis and on-line analytical processing;
- Develop a case for the business value of visualization of data; and
- Apply data mining tools for discovery of hidden relationships and patterns in business data.

13.3 OPERATIONAL AND INFORMATIONAL DATABASES

There are several reasons that have compelled information system designers to accept that even the latest technology cannot be optimized to support both operational and informational (or analytical) processing in a cost effective manner. The reasons may be like following -

- The data required is different
- The information required is different
- The technology required is different
- The user community is different
- The access pattern is different
- The processing characteristics are different
- The system loading pattern is different and further more.

High level business objectives provide the framework for competitiveness and growth to an enterprise in the long run. To set realistically achievable objectives and monitor the company's performance against them, managers need information on *subjects* like customers (e.g. their needs and preferences), products (e.g. the company's offering vis-à-vis competitors), suppliers (e.g. their production lead times) and distributors (e.g. location). It is also beneficial to *aggregate* measures like quantities of products and sales values across various *dimensions* like geographical regions and time, and *compare* them with yard-sticks like budgets, forecasts, previous year's actuals and industry averages. In today's world surrounded by IT and its applications it is more imperative as they may be considered as the basic requirements for on-line businesses. The differences between informational systems satisfying needs of these types and traditional operational systems are summarized in the Table 13.1.

Table 13.1: Operational and Decision Support Data

PRIMITIVE DATA / OPERATIONAL DATA	DERIVED DATA / DSS DATA
application oriented detailed accurate, as of the moment of access serves the clerical community can be updated run repetitively	subject oriented summarized, otherwise refined represents values over time, snapshots serves the managerial community is not updated

requirements for processing understood a priori compatible with the SDLC performance sensitive accesses a unit at a time transaction driven control of update a major concern in terms of ownership high availability managed in its entirety nonredundancy static structure; variable contents small amount of data used in a process supports day-to-day operations high probability of access	run heuristically requirements for processing not understood a priori completely different life cycle performance relaxed accessed a set at a time analysis driven control of update no issue relaxed availability managed by subsets redundancy is a fact of life flexible structure large amount of data used in a process supports managerial needs low, modest probability of access
--	---

Information systems have evolved from the master-file / transaction file and sequential processing of the 1960s to DBMS based random access processing of the 1970s to high-performance transaction processing during the 80s to decision support systems built around the existing operational, informational, and virtual databases. The traditional IT departments with limited resources, when faced with the need to provide strategic information using data in operational databases find themselves inadequately equipped to respond. Some of the difficulties faced are:

- IT receives too many ad hoc requests, resulting in a large overload. With limited resources, IT is unable to respond to the numerous requests in a timely fashion.
- Requests are not only too numerous, they also keep changing all the time. The users need more reports to expand and understand the earlier reports.
- The users find that they get into the spiral of asking for more and more supplementary reports, so they sometimes adapt by asking for every possible combination, which only increases the IT load even further.
- The users have to depend on IT to provide the information. They are not able to access the information themselves interactively, most of the time.
- The information environment, ideally suited for strategic decision-making, has to be very flexible and conducive for analysis. IT has been unable to provide such an environment.

The solution to the problems lies in the paradigm – the *data warehouse*. It is an *environment* that is specifically designed to hold data required to support complex analysis, and detection of patterns, trends and deviations. The characteristics of such an environment can be summarized as follows:

- Provides an integrated and total view of the enterprise.
- Make the enterprise's current and historical information easily available for decision-making.

- Makes decision-support transactions possible without hindering operational systems.
- Renders the organization's information consistent.
- Presents a flexible and interactive source of strategic information.

13.4 THE DATA WAREHOUSE

We have understood the need and importance of data warehouse to facilitate strategic decision-making. The concept of a data warehouse given by Bill Inmon, the father of data warehousing, is depicted in Figure 13.1.

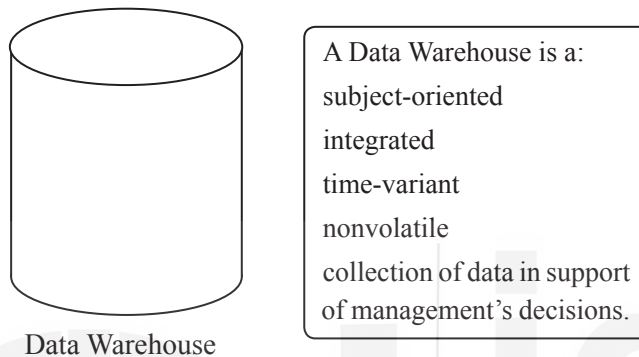


Figure 13.1: What is a Data Warehouse?

The defining characteristics of a data warehouse are:

Subject-orientation: Data warehouse data of an organization are arranged and optimized to provide answers to questions coming from diverse functional area within a company. Therefore, the data warehouse contains data organized and summarized by meta data and keywords i.e. by topic, such as sales, marketing, finance, distribution, transportation etc. For each one of these topics the data warehouse contains specific subjects of interest - products, customers, departments, regions, promotions, and so on. Note that this form of data organization is quite different from the more functional or process-oriented organization of typical transaction systems.

Time-variancy: We have already noted that the DSS data include a time element (see Table 13.1). In contrast to the operational data, which focus on current transactions, the warehouse data represent the flow of data through time. The data warehouse can even contain projected data generated through statistical and other models.

Non-volatility: Once data enter the data warehouse they are never removed. Because the data in the data warehouse represent the company's entire history, the operational data representing the near-term history, are always added to it. The near-term history information is gathered and stored by data collecting IT tools. One such example by IBM is OMEGAMON XE for their DB2 RDBMS. Because data are never deleted and new data are always added, the data warehouse is always growing. That is why the DSS DBMS must be able to support storage in terms of multi-gigabyte and even multi-terabyte database and multiprocessor hardware. However, based on the organizational policy, an organization may take decision to remove old data from the warehouse after a period of ,say, 10 years. It may also be

possible to offload such data to another secondary storage for reference in future.

Integration: The data warehouse is a centralized, consolidated database that integrates data derived from the entire organization. Thus the data warehouse consolidates data from multiple and diverse sources with diverse formats. Data integration implies a well-organized effort to define and standardize all data elements. This integration effort can be time-consuming but, once accomplished, it provides a unified view of the overall organizational situation. Data integration enhances decision-making and helps managers to better understand the company's operations. This understanding can be translated into recognition of strategic business opportunities.

Table 13.2 summarizes the differences between the data in a data warehouse and that in an operational database.

Table 13.2: Operational Data and Data Warehouse Data

Characteristic	Operational Database Data	Data Warehouse Data
Integrated	Similar data can have different representations or meanings. For example, telephone numbers may be stored as 033-29-70701 or as 0332970701, and a given condition may be labeled as T/F or 0/1 or Y/N. A sales value may be shown in thousands or in millions.	Provide an unified view of all data elements with a common definition and representation for all business units.
Subject-oriented	Data are stored with a functional, or process, orientation. For example, data may be stored for invoices, payments, credit amounts, and so on.	Data are stored with a subject orientation that facilitates, multiple views of the data and facilitates decision making. For example, sales may be recorded by product, by division, by manager, or by region.
Time-variant	Data are recorded as current transactions. For example, the sales data may be the sale of a product on a given date, such as Rs. 59,342.78 on 08-NOV-2021.	Data are recorded with a historical perspective in mind. Therefore, a time dimension is added to facilitate data analysis and various time comparisons.
Non-volatile	Data updates are frequent and common. For example, an inventory amount changes with each sale. Therefore, the data environment is fluid.	Data cannot be changed. Data are only added periodically from historical systems. Once the data are properly stored, no changes are allowed. Therefore the data environment is relatively static.

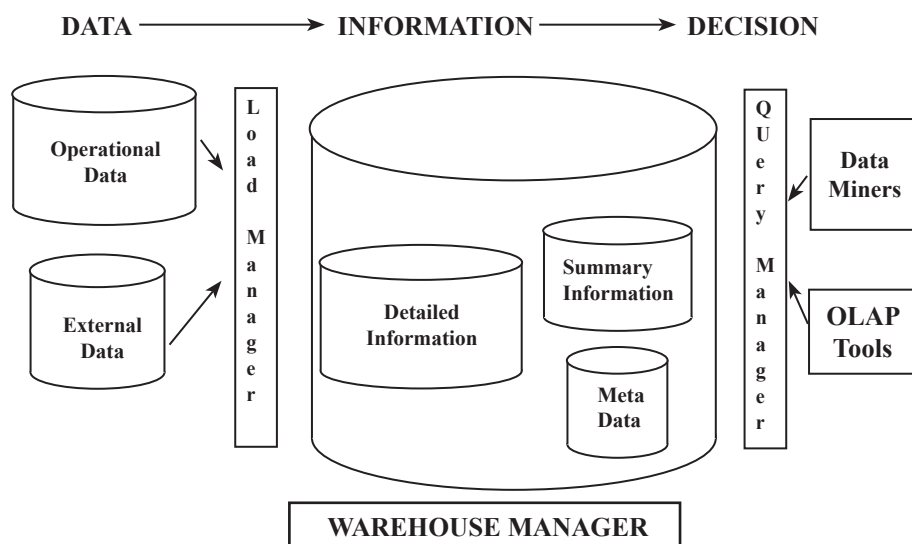


Figure 13.2: Depicts the architecture of a data warehouse

THE LOAD MANAGER

Data flows into the data warehouse through the *load manager*. The data is mostly extracted from the operational database(s) and other internal sources (like archived historical data), and supplemented by data imported from external sources. Externally sourced data can greatly enhance the value of information generated from a data warehouse. For example a gas pipeline operator in western countries like UK, USA, CANADA etc may use weather forecast data from the country's Meteorological Office on a regular basis to determine demand for gas (the main source of energy used for heating homes and offices) in various areas of the country. The weather data may be fed into a *model* incorporating several other factors (e.g. day of the week, internal data about customers' usage patterns, demographic and economic profile data, alternate sources of energy, types of buildings in the area) to arrive at a demand forecast. Types of data from external sources that may be included in data warehouse are: financial indicators and statistics of the industry, market share data of competitors, demographic data, weather data, credit worthiness data, readership / viewer survey data for advertising media, specially commissioned surveys and so on. External data is usually obtained from commercial database services or government agencies (e.g. Met Office, census agency, industry associations, stock exchanges, local government statistics service). The data from such diverse sources will obviously be in different incompatible formats and will be distributed through various media. Some of them may be available on a downloadable format from virtual data bases on the Internet; others may be distributed on secondary storages while some may only be available on printed media. Some data may be available for free but most data (particularly when used for commercial purposes) have to be purchased.

The load manager primarily performs what is termed an *extract-transform-load (ETL)* operation.

- Data Extraction
- Data Transformation
- Data Loading

Data Extraction: This function has to deal with numerous data sources. Appropriate techniques have to be employed for each data source. Source data may be from different source machines in diverse data formats. Part of the source data may be in relational database systems. Some data may be on other legacy network and hierarchical data models. Many data sources may still be in flat files. There may also be the need to include data from spreadsheets and local departmental data sets. Data extraction can become quite a complex operation at times.

Various tools are available on the market for data extraction. Use of outside tools may be considered suitable for certain data sources. For the other data sources, in-house programs may be also needed to be developed to do the data extraction. Purchasing outside tools may entail high initial costs. In-house programs, on the other hand, may mean ongoing costs for development and maintenance.

After extraction, the data needs to be kept somewhere for further preparation. Sometimes the extraction function is performed in the legacy platform itself if that approach suits the designed framework. More frequently, data warehouse implementation teams extract the source data into a separate physical environment from which moving the data into the data warehouse would be easier. In the separate environment, the source data may be extracted into a group of flat files, or an intermediate relational database, or a combination of both. This physical environment is called the *data-staging* area.

Data Transformation: In every system implementation, data conversion is an important function. For example, when implementing an operational system such as a magazine subscription application, the database has to be initially populated with data from the existing system records. The conversion may either be from a manual system or from a file-oriented system to a modern system supported with relational database or object oriented tables. In either case, the data will need to be converted from the existing systems. So, what is so different for a data warehouse? Why is data transformation for a data warehouse more involved than that for an operational system?

As already discussed, data for a data warehouse comes from many disparate sources. If data extraction for a data warehouse poses great challenges, data transformation presents even greater challenges. Another factor in the data warehouse is that the data feed is not just an initial one-time load. The ongoing changes will have to continue to be picked up from the source systems. Any transformation tasks are set up for the initial load will have to be adapted for the ongoing revisions as well.

A number of individual tasks are performed as part of data transformation. First, the data extracted from each source is cleaned. Cleaning may be correction of misspellings, or may include resolutions of conflicts between state codes and pin codes in the source data, or may deal with providing default values for missing data elements, or elimination of duplicates when the same data is brought in from multiple source systems.

Standardization of data elements forms a large part of data transformation. The data types and field lengths for same data elements retrieved from the

various sources need to be standardized. Semantic standardization is another major task. *Synonyms* and *homonyms* have to be resolved. Resolution of synonyms is required when two or more terms from different source systems mean the same thing. On the other hand, when a single term means many different things in different source systems, resolution of homonyms have to be performed.

Data transformation involves many forms of combining pieces of data from the different sources. In some cases, data from a single source record or related data element from many source records are combined. In other situations, data transformation may also involve purging source data that is not useful and/or separating out source records into new combinations. During data transformation sorting and merging of data takes place on a large scale in the data staging area.

In many cases, the keys chosen for the operational systems are field values with built-in meanings. For example, the product key value may be a combination of characters indicating the product category, the code of the warehouse where the product is stored, and some code to show the production batch. Primary keys in the data warehouse cannot have built-in meanings. Therefore, data transformation also includes the assignment of surrogate keys derived from the source system primary keys.

A grocery chain point-of-sale operational system keeps the unit sales and revenue amounts by individual transactions at the checkout counter at each store. But in the data warehouse, it may not be necessary to keep the data at this detailed level. It may be more appropriate to summarize the totals by product at each store for a given day and keep the summary totals of the sale units and revenue in the data warehouse's storage. In such cases, the data transformation function would include such summarization processing.

The end result of the data transformation function is a collection of integrated data that is cleaned, standardized, and summarized. Now the data is ready to be loaded for populating the data warehouse.

Data Loading: Two distinct groups of tasks form the data loading function. After completion of the design and construction of the data warehouse, when it goes live for the first time, the initial loading of data is done. The initial load moves large volumes of data which may take substantial amount of time, but it is a one-time effort. As the data warehouse starts functioning, extraction of additions (and changes) to the source data continues on an ongoing basis, together with the transformation and loading operations.

THE QUERY MANAGER

The *query manager* provides an interface between the data warehouse and its users. It performs tasks like directing the queries to the appropriate tables, generating views on an ad-hoc basis if required, monitoring the effectiveness of indexes and summary data, and query scheduling.

DATA WAREHOUSE DESIGN CONSIDERATIONS

The key considerations involved in the design of a data warehouse are:

- Time Span
- Granularity

- Dimensionality
- Aggregations
- Partitioning

Time Span: Operational data represent current (atomic) transactions. Such transactions might define a purchase order, a sales invoice, an inventory movement, and so on. In short, operational data cover a short time frame. In contrast, data warehouse data tend to cover a longer time frame. Managers are seldom interested in a specific sales invoice to customer X; rather they tend to focus on sales generated during the last month, the last year, or the last five years. Rather than concern themselves with a single customer purchase, they might be interested in the buying pattern of such a customer or groups of customers. In short, data warehouse data tend to be historic in nature. That is, the data warehouse data represent company transactions up to a given point in time, yesterday, last week, last month, and the like. The time period for which data is held in the data warehouse is determined by the data analysis requirements of the users of the data warehouse. These needs, in turn, arise from the changes in the business environment that a particular organization needs to monitor, in its effort to stay ahead of its competitors. Since, a data warehouse's size depends on the span of time for which data is stored, the time span covered by the data warehouse is an important design consideration. If, for example, the environment changes rapidly, the data required for analysis would relate more often to the *recent past*, rather than that over several years or decades. In that case the designers of the data warehouse need to consider whether or not the cost incurred in holding data for indefinitely long time spans would be worthwhile.

Granularity: According to Inmon, the single most important design aspect of a data warehouse is the decision on granularity. It refers to the level of detail or summarization available in units of data in the data warehouse. The more detail there is, the lower the level of granularity. The less detail there is, the higher the level of granularity.

Operational data represent specific transactions that occur at a given time, such as customer purchase of product X in store A. Thus, granularity is taken for granted to be of the lowest level, in operational systems. Data warehouse data must be presented at different levels of aggregation, from highly summarized to near atomic. This requirement is based on the fact that managers at different levels in the organization require data with different levels of aggregation. It is also possible that a single problem requires data with different summarization levels. For example, if a manager must analyze sales by region, (s)he must be able to access data showing the sales by region, by city within the region, by store within the city within the region, and so on. In this case, the manager requires summarized data to compare the regions, but (s)he also needs data in a structure that enables him or her to decompose (drill down) the data into more atomic components (that is, data at lower levels of aggregation). For example, it is necessary to be able to drill down to the stores within the region in order to compare store performance by region. Granularity level in a data warehouse cannot, therefore, be assumed.

The decision on granularity level profoundly affects both the volume of data that resides in the data warehouse, and the type of query that can be answered. A trade off exists between the volume of data in the data warehouse and the level of detail of queries (see Figure 13.3 below). Some data warehouses are designed to support *dual granularity*. In such environments some data (usually the most recent) is held at a relatively low level of granularity, while the rest is held in more summarized form (i.e. at a higher granularity level). This enables detailed analysis at the same time allows reduction of data volume.

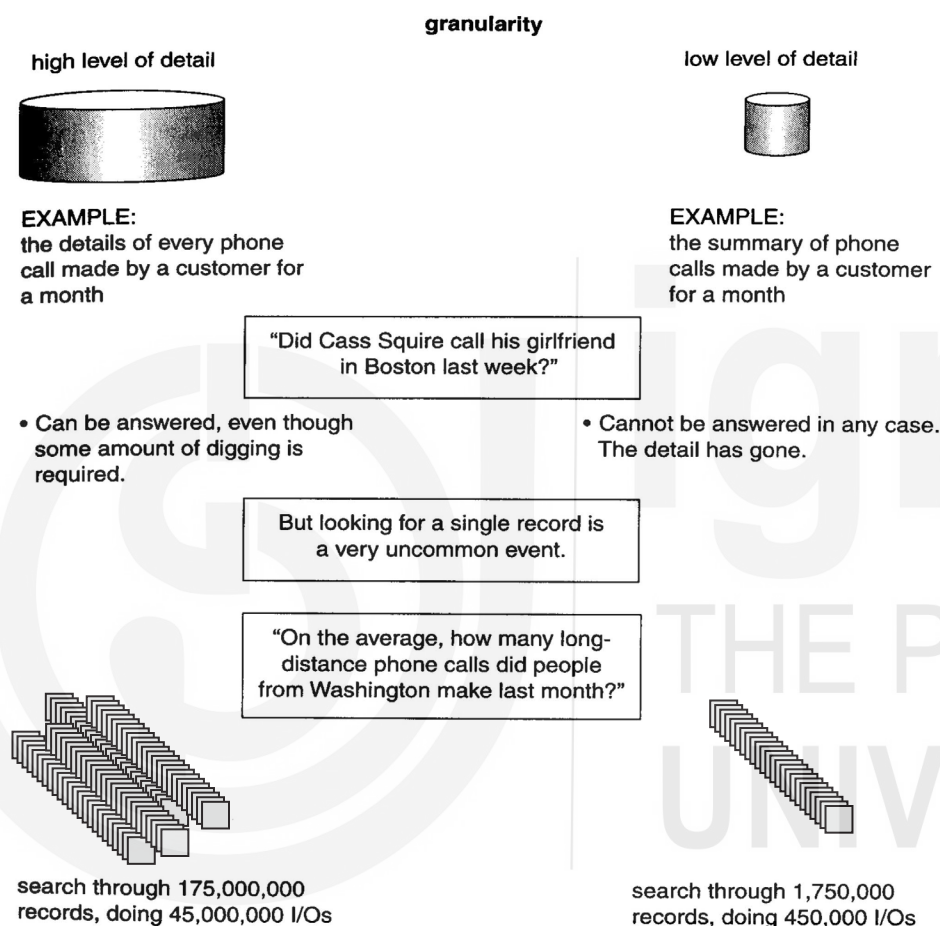


Figure 13.3: Granularity of Data Warehouse Data

Dimensionality This is probably the most distinguishing characteristic of a data warehouse. From the data analyst’s point of view, the data is always related in many different ways. For example, when we analyze product sales by a customer during a given time span, we are likely to ask how many widgets of type X were sold to customer Y during the last six months. In fact, the question tends to expand quickly to include many different data *dimensions*. For instance, we might want know how the product X fared relative to product Z during the past six months, by region, state, city, store, and customer (or sales of various products by quarters by country, as shown in Figure 13.4 below). In this case, both place and time are part of the picture. In general, data analysis tends to include many data dimensions, producing a multidimensional view of the data.:

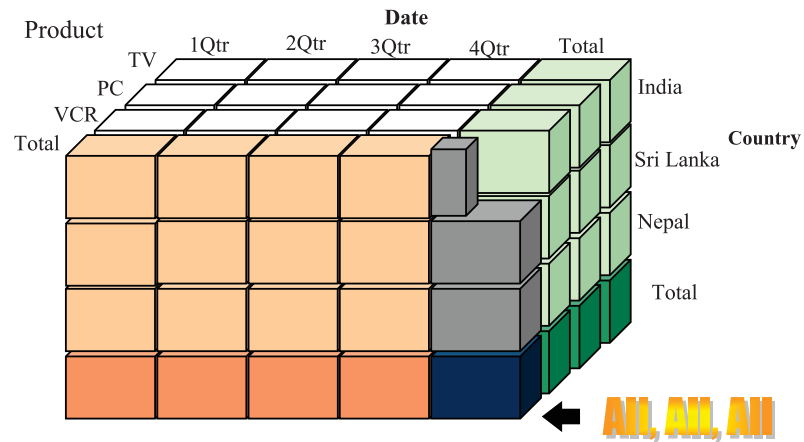


Figure 13.4: Dimensionality of Data Warehouse Data

The data model used for modeling data warehouses is known as the *dimensional model*. The numerical measurements related to the business (like sales volumes) are stored in *fact tables*. The descriptions of the dimensions are stored in *dimension tables*. We will discuss this in more detail in later sections. The number and types of dimensions and facts that are to be stored in a data warehouse is a very important design decision, and (much like the decision on granularity) affects both the data volume and the types of analysis that can be supported.

Aggregations: We have seen how data analysis queries directed at data warehouses involve dimensions. Another very common type of query directed at data warehouses involves sums of values along the different dimensions. For example: what is the total sales volume of VCR during the past 4 quarters? Answering this query using the fact and dimension tables would involve summing up the individual sales volume figures over the 4 quarters and the 3 counties. In real situations, similar queries might involve retrieving and summing hundreds or thousands of individual values. To avoid excessive processing load on the data warehouse arising from frequently asked queries of this type, it is often decided, at design time, to store some *pre-calculated aggregations*, along with the base facts, in the data warehouse (as illustrated in Figure 13.4 above). This decision affects the data volume and performance of certain types of queries.

Partitioning: One of the essences of the data warehouse is flexible data storage, management, and access. When data resides in large physical units, among other things it cannot be:

- indexed easily
- sequentially scanned, if needed
- restructured easily
- backed up conveniently
- recovered easily
- monitored easily and so on.

In short, having a big mass of data defeats much of the purpose of the data warehouse. The purpose of *partitioning* is to break the data into smaller (more manageable) physical units of storage. The criteria used for dividing

the data can be: date, line of business / product category, geography / location, organizational / administrative unit, or any combination of these.

Activity A

What is the difference between a database and a data warehouse? Take a database from any organization and try to convert it into a data warehouse. What are the visible advantages that you can make out?

.....

.....

.....

.....

13.5 DATA WAREHOUSE SCHEMA

One of the key questions to be answered by the database designer is: How can we design a database that allows unknown queries to be performant? This question encapsulates the differences between designing for a data warehouse and designing for an operational data base and information system. In a data warehouse one designs to support the *business process* rather than specific query requirements. In order to achieve this, the designer must understand the way in which the information within the data warehouse will be used.

In general, the queries directed at a data warehouse tend to ask questions about some essential fact, analyzed in different ways. For example reporting on:

- The average number of light bulbs sold per store over the past month
- The top ten most viewed TV programs, from a TV channel or channels, during the past week
- Top spending customers over the past quarter
- Customers with average credit card balances more than Rs.50,000 during the past year.

Each of these queries has one thing in common: they are all based on factual data. The content and presentation of the results may differ between examples, but the factual and transactional nature of the underlying data is the same.

Table 13.3: Fact Tables and Attributes

Requirement	Fact	Attributes
Sales of light bulbs	Transaction in a period	Quantity sold Product identifier Store identifier Data and time Revenue achieved
TV Channels programs	Channel pay-per-view transaction	Customer identifier TV Channel watched Program watched Data and time Duration Household identifier

Customer spend	Loyalty card transaction	Customer Identifier Store identifier Transaction value Data and time
Customer account balance Account transactions		Customer identifier Account number Type of transaction Amount of transaction Destination account number

Fact data possesses some characteristics that allow the underlying information in the database to be structured. Facts are transactions that have occurred at some point in the past, and are unlikely to change in the future. Facts can be analyzed in different ways by cross-referencing the facts with different reference information. For example, we can look at sales by store, sales by region, or sales by product. In a data warehouse facts also tend to have few attributes, because there are no operational data overheads. For each of the examples described, the attributes of the fact could be as listed in Table 13.3 above.

One of the major technical challenges within the design of a data warehouse is to structure a solution that will be effective for a reasonable period of time (at least three to five years). This implies that the data should not have to be restructured when the business changes or the query profiles change. This is an important point, because in more traditional applications it is not uncommon to restructure the underlying data in order to address query performance issues.

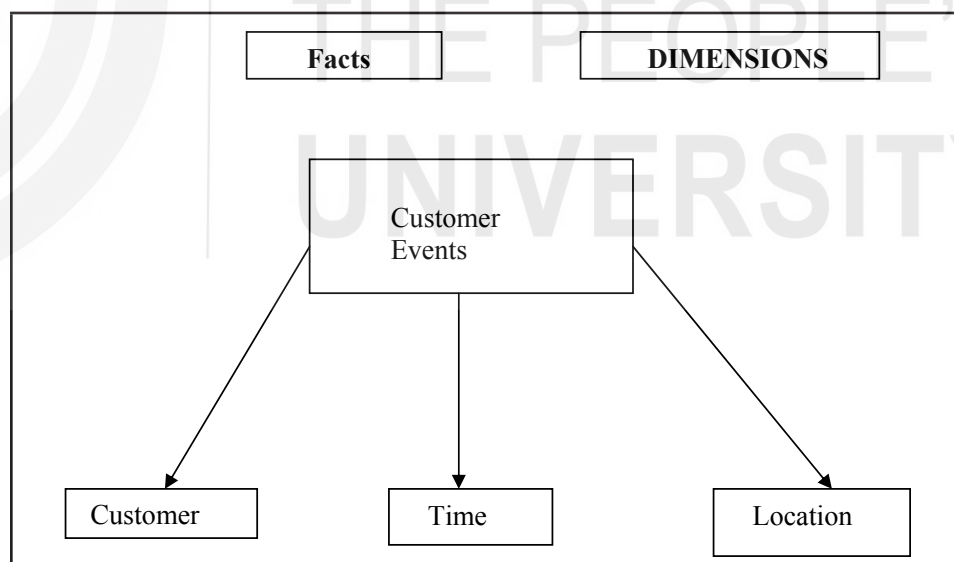


Figure 13.5: Star Schema

The inherent advantage of factual transactions is that their content is unlikely to change, regardless of how it is analyzed. Also, the majority of the data volume of a data warehouse comes from the factual information. It is therefore possible to treat fact data as read-only data, and the reference data (used to interpret the fact data) as data that is liable to change over time. Thus, if reference data needs to change, the voluminous fact data would not have to be changed or restructured.

Star schemas (so called due to their ‘star like’ appearance) are physical database structures that store the factual data in the ‘center’, surrounded by the reference (or dimension) data (see Figure 13.5 above).

The dimension tables should be relatively small in comparison to the size of the data warehouse, so that restructuring costs are small as long as the keys to the fact tables are not changed. In star schema arrangements, the reference information is often de-normalized to a single table to speed up query performance. The redundancy overheads are acceptable, as the sizes involved are small and even the reference information changes infrequently.

The dimensional information, represented by the reference data is often organized in form of *concept hierarchies* that are used for analysis, and designing data warehouse aggregations. A typical concept hierarchy for a retail chain is depicted in Figure 13.6 below.

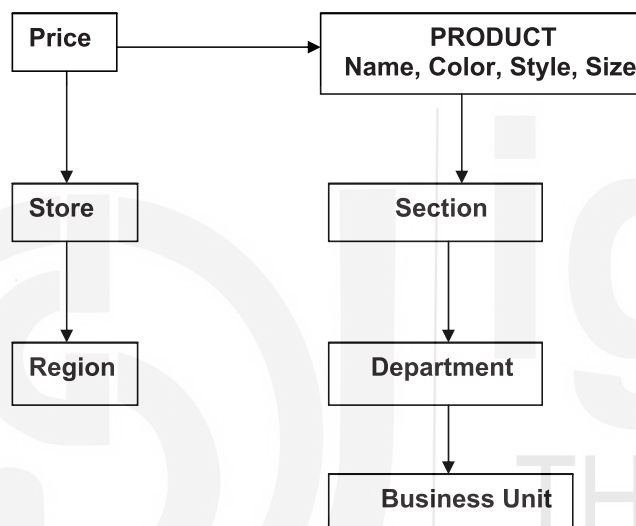


Figure 13.6: Concept Hierarchies

The number of concept hierarchies that can be defined on any given dimension is by no means restricted to one. There may be several arbitrary concept hierarchies in use in any organization to enable data analysis from various angles. Figure 13.7 below shows how the ‘time’ dimension in a retail data warehouse (represented by the day / date reference data) may be organized into several concept hierarchies or *groupings* (which are single level concept hierarchies – ‘Easter’ and ‘Summer’ in Figure 13.7).

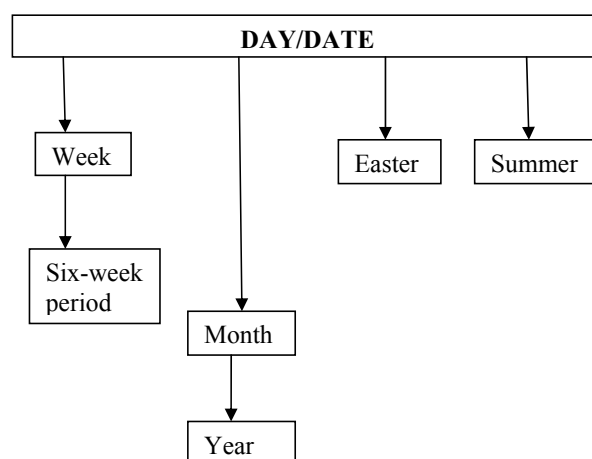


Figure 13.7: Multiple Hierarchies

When the concept hierarchies and groupings are incorporated into a star schema diagram (like Figure 13.8 below), the appearance resembles a 'snowflake'. Hence, schemas of this type are called *Snowflake schemas*.

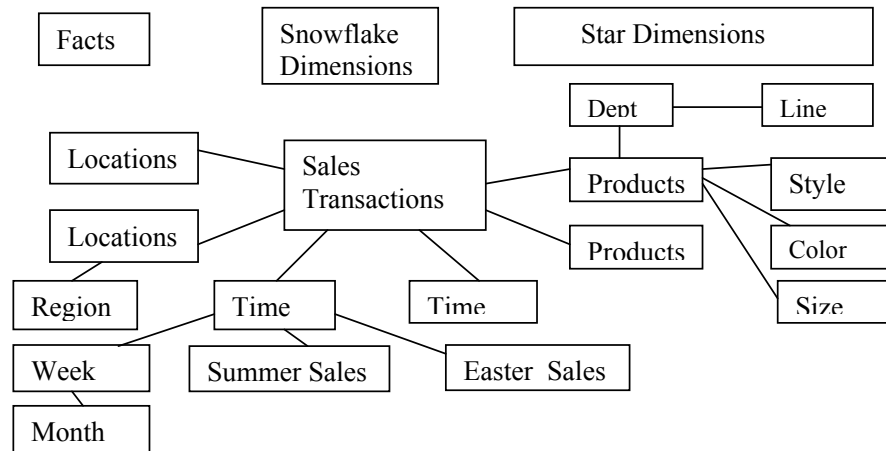


Figure 13.8: Snowflake Schema

13.6 METADATA

Metadata in a data warehouse is similar to the data dictionary in the context of a database. It stores data about data in the data warehouse.

Types of Metadata

Metadata in a data warehouse fall into three major categories:

- Operational Metadata
- Extraction and Transformation Metadata
- End-User Metadata

Operational Metadata: As already discussed, data for the data warehouse comes from several operational systems of the enterprise. These source systems contain different data structures. The data elements selected for the data warehouse have various field lengths and data types. Selecting data from different source files, and loading it into the data warehouse, requires splitting of records, combining parts of records from different source files, and dealing with multiple coding schemes and field lengths. When information is delivered to the end-users, it is essential to be able relate back to the original source data sets. Operational metadata contain all of this information about the operational data sources that allow us to trace back to the original source.

Extraction and Transformation Metadata: Extraction and transformation metadata contain data about the extraction of data from the source systems, namely, the extraction frequencies, extraction methods, and business rules for the data extraction. Also, this category of metadata contains information about all the data transformations that take place in the data staging area.

End-User Metadata. The end-user metadata is the navigational map of the data warehouse. It enables the end-users to find information from the data warehouse. The end-user metadata allows the end-users to use their own business terminology and look for information in those ways in which they normally think of the business.

Special Significance

Why is metadata especially important in a data warehouse?

- First, it acts as the glue that connects all parts of the data warehouse.
- Next, it provides information about the contents and structure to the developers.
- Finally, it opens the door to the end-users and makes the contents recognizable in their own terms.

Metadata Requirements

According to Inmon, a new user approaching a data warehouse wants to know

- What tables, attributes, and keys does the data warehouse contain?
- From where did each set of data come?
- What transformation logic was applied in loading the data?
- How has the data changed over time?
- What aliases exist, and how are they related to each other?
- What are the cross-references between technical and business terms?
- How often does the data get reloaded?
- How much data is there? This helps end-users to avoid submitting unrealistic queries. Given some means of determining the size of tables, technical staff can tell the end users, “You can do what you like with 15,000 rows, but if it turns out to be more then back off and ask for help!”

Metadata requirements of various classes of users are summarized in Table 13.4 below:

Table 13.4: Uses of Metadata

	IT Professionals	Power Users	Casual Users
Analysis and Discovery	Database Tables, Columns, Server Platforms.	Databases, Tables, Columns	List of Predefined Queries and Reports, Business views.
Meaning of Data	Data structures, Data Definitions, Data Mapping, Cleansing Functions, Transformation Rules	Business Terms, Data Definitions, Data Mapping, Cleansing Functions, Transformation Rules	Business Terms, Data Definitions, Filters, Data Sources, Conversion, Data Owners
Information Access	Program Code in SQL, 3GL, 4GL, Front-end Applications, Security	Query Toolsets, Database Access for Complex Analysis	Authorization Requests, Information Retrieval into Desktop Applications such as Spreadsheets.

Metadata Components

Warehouse metadata is not very different in kind from ordinary database metadata, although it is versioned in order to permit historical analysis. Prism gives the following breakdown of warehouse metadata in its Tech Topic, “Metadata in the Data Warehouse:”

Mapping

The mapping information records how data from operational sources is transformed on its way into the warehouse. Typical contents are:

- Identification of source fields
- Simple attribute-to-attribute mapping
- Attribute conversions
- Physical characteristic conversions
- Encoding/reference table conversions
- Naming changes
- Key changes
- Defaults
- Logic to choose from among multiple sources
- Algorithmic changes

Extract History

Whenever historical information is analyzed, meticulous update records have to be kept. The metadata history is a good place to start any time-based report, because the analyst has to know when the rules changed in order to apply the right rules to the right data. If, for example, sales territories were remapped in 1991, results from before that date may not be directly comparable with more recent results.

Miscellaneous

- Aliases can make the warehouses much more use-friendly by allowing a table to be queried by “Widgets produced by each factory” rather than “MF-STATS.” Aliases also come in useful when different departments want to use their own names to refer to the same underlying data. Obviously, though, aliases can also cause a great deal of confusion if they are not carefully tracked.
- Often, parts of the same data warehouse may be in different stages of development. Status information can be used to keep track of this: for instance, tables might be classified “in-design,” “in-test,” “inactive,” or “active.”
- Volumetric information lets users know how much data they are dealing with, so that they can have some idea about how much their queries will cost in terms of time and computational resources. Volumetrics could usefully include such information as number of rows, growth rate, usage characteristics, indexing, and byte specifications.
- It is also useful to publish the criteria and time scales for purging old data.

Summarization and Aggregation Algorithms

As discussed above, a typical data warehouse contains lightly and heavily summarized data, and aggregations as well as full detailed records. The algorithms for summarizing (and aggregating) the detail data are obviously of interest to anyone who takes responsibility for interpreting the meaning of the summaries. This metadata can also save time by making it easier to decide which level of summarization is most appropriate for a given purpose.

Relationship Artifacts and History

Data warehouses implement relationships in a different way from production databases. Metadata pertaining to related tables, constraints, and cardinality are maintained, together with text descriptions and ownership records. This information and the history of changes to it can be useful to analysts.

Ownership/Stewardship

Operational databases are often owned by particular departments or business groups. In an enterprise data warehouse, however, all data is stored in a common format and accessible to all. This makes it necessary to identify the originator of each set of data, so that inquiries and corrections can be made by the proper group. It is useful to distinguish between *ownership* of data in the operational environment and *stewardship* in the data warehouse.

Access Patterns

It is desirable to record patterns of access to the warehouse in order to optimize and tune performance. Less frequently used data can be migrated to cheaper storage media, and various methods can be used to accelerate access to the data that is most in demand. Most databases do a good job of hiding such physical details, but specialized performance analysis tools are usually available. Some general-purpose tools, such as Information Builders' SiteAnalyzer, also are available.

Reference Tables / Encoded Data

Reference data is stored in an external table (see discussion on Star Schema above) and contains commonly used translations of encoded values. The contents of these tables must be stored in order to guarantee the ability to recover the original un-encoded data, together with effective from and effective to dates.

Data Model – Design Reference

Building a data warehouse without first constructing a data model is very difficult and frustrating. When a data model is used, metadata describing the mapping between the data model and the physical design should be stored. This allows all ambiguities or uncertainties to be resolved.

From the point of view of the Query Manager (see Figure 13.2 above) of the data warehouse, the *Metadata Repository* can be perceived to have three logical layers: the *Information Navigator*, the *Business Metadata*, and the *Technical Metadata*. Figure 13.9 below illustrates this concept. The query manager accesses the metadata through the Information Navigator layer which is the topmost layer of the metadata repository. The higher layers, in

turn, access more detailed metadata components resident in the lower layers whenever required.

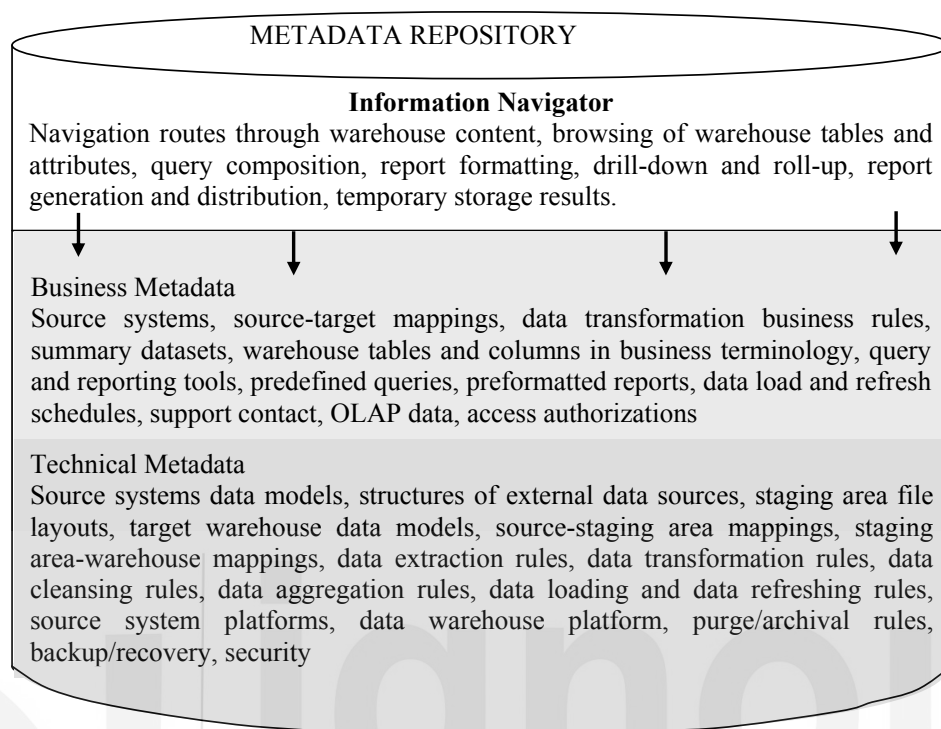


Figure 13.9: The Metadata Repository

13.7 DATA WAREHOUSE AND THE WEB

Professor Peter Drucker, the senior guru of management practice, has admonished IT executives to look outside their enterprises for information. He remarked that the single biggest challenge is to organize outside data because change occurs from the outside. He predicted that the obsession with internal data would lead to being blindsided by external forces.

The majority of data warehousing efforts result in an enterprise focusing inward; however, the enterprise should be keenly alert to its externalities. As markets become turbulent, an enterprise must know more about its customers, suppliers, competitors, government agencies, and many other external factors. The changes that take place in the external environment, ultimately, get reflected in the internal data (and would be detected by the various data analysis tools discussed in the later sections), but by then it may be too late for the enterprise – proactive action is always better than reacting to external changes after the effects are felt. The conclusion is that the information from internal systems must be enhanced with external information. The synergism of the combination creates the greatest business benefits.

The importance of external data and the challenges faced in integrating external data with internally sourced data was discussed in the Section on Load Manager above. In the same section, it was mentioned that, some externally sourced data (particularly time sensitive data), is often distributed through the internet.

RELIABILITY OF WEB CONTENT

Many question the reliability of web content, as they should. However, few analyze the reliability issue to any depth. The Web is a global bulletin board on which both the wise and foolish have equal space. Acquiring content from the Web should not reflect positively or negatively on its quality.

Consider the following situation: If you hear, “Buy IBM stock because it will double over the next month,” your reaction should depend on who made that statement and in what context. Was it a random conversation overheard on the subway, a chat with a friend over dinner, or a phone call from a trusted financial advisor? The *context* should also be considered when judging the reliability of Web content.

Think of Web resources in terms of quality and coverage, as shown in Figure 13.10 below.

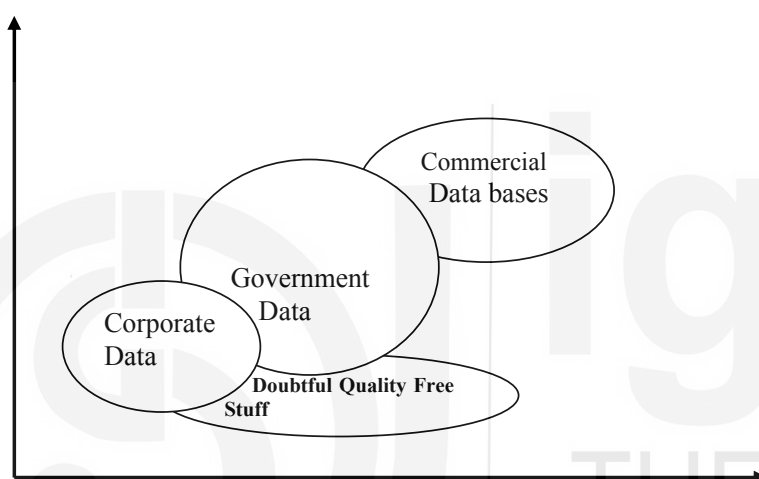


Figure 13.10: Web-based Information Sources

In the Figure 13.10, toward the top are information resources of high quality (accuracy, currency, and validity), and resources toward the right have a wide coverage (scope, variety, and diversity). The interesting aspect of the web is that information resources occupy all quadrants.

In the upper center, the commercial online database vendors traditionally have supplied business with high-quality information about numerous topics. However, the complexities of using these services and the infrequent update cycles have limited their usefulness.

More to the left, governmental databases have become tremendously useful in recent years. Earlier, public information was often available only by spending many hours of manual labour at libraries or government offices. Now facilities on internet are available to provide valuable and up-to-date data via the Web and virtual data bases. Public, organizations, governments, business houses, educational institutions and other industries realize the importance of data also with regard to ‘right to inform’ and hence up-to-date data and information is required to be maintained as obligatory requirements of the user community.

At the left, corporate Web sites often contain vast amounts of useful information in white papers, product demos, audios/videos, and press

releases, eliminating the necessity to attend trade exhibits to learn the “latest and greatest” in a market place.

Finally, the “doubtful-quality free” content occupies the lower half of the figure. Its value is not in the quality of any specific item but in its constantly changing diversity. Combined with the other Web resources, the doubtful-quality free content acts as a wide-angle lens to avoid tunnel vision of the market place.

WEB FARMING

Like operational systems, the *Web farming* system provides input to the data warehouse. The result is to disseminate the refined information about specific business subjects to the enterprise sourced from the Web.

The primary source of content for the Web farming system is the Web because of its external perspectives on the business of the enterprise. As a content source, the Web can be supplemented (but not replaced) by the intranet web of the enterprise. This content is typically in the format of internal Web sites, word processing documents, spreadsheets, and e-mail messages. However, the content from the intranet is usually limited to internal information about the enterprise, thus negating an important aspect of Web farming.

Most information acquired by the Web farming system will not be in a form suitable for the data warehouse. Also, as discussed above, the source and quality of the content need to be judged. In any case, the information must be *refined* before loading into the warehouse. However, even in its unrefined state, the information obtained from the Web, through Web farming, could be highly valuable to the enterprise. The capability to directly disseminate this information may be required via textual message alerts or “What’s New” bulletins.

REFINING INFORMATION

When a data warehouse is first implemented within an enterprise, a detailed analysis and reengineering of data from operational systems is required (see Section on Load Manager above). The same is true for Web farming. Before Web content can be loaded into a warehouse, the information must be refined.

The processes of refining information consists of four steps: *discovery*, *acquisition*, *structuring*, and *dissemination*.

Discovery is the exploration of available Web resources to find those items that relate to specific topics. Discovery involves considerable detective work far beyond searching generic directory services, such as Google!, or indexing services, such as Alta Vista. Wikipedia is another search service. There are many more known as search engines to dig out data and information from the world wide web. The discovery activity must be a continuous process because data sources are continually appearing and disappearing from the Web.

Acquisition is the collection and maintenance of content identified by its source. The main goal of acquisition is to maintain the historical context, such as in Wikipedia, so that you can analyze content in the context of its

past. A mechanism to efficiently use human judgement in the validation of content is another key requirement.

Structuring is the analysis and transformation of content into a more useful format and into a more meaningful structure. The formats can be Web pages, spreadsheets, word processing documents, and database tables. As we move toward loading data into a warehouse, the structures must be compatible with the star-schema design and with key identifier values.

Dissemination is the packaging and delivery of information to the appropriate consumers, either directly or through a data warehouse. A range of dissemination mechanisms is required, from predetermined schedules to ad hoc queries. Newer technologies, such as information brokering and preference matching, may be desirable.

13.8 ON-LINE ANALYTICAL PROCESSING (OLAP)

In the previous sections we have discussed the need for building an Enterprise Data Warehouse, followed by discussions on its structure and ways and means of populating it with data. The data warehouse serves as the ‘memory’ of the enterprise. But, memory is of little use without intelligence. The tools that analyze data provide the ‘intelligence’. In this section and the ones that follow, we discuss some of the tools available for accessing the data warehouse and transforming the data in it into information useful from the point of view of the business or *Business Intelligence*.

The term, On-line Analytical Processing (OLAP) was coined by E.F. Codd in 1993 to refer to a type of application that allows a user to interactively analyze data. Online Analytical Processing (OLAP) is a method of analyzing data in a multi-dimensional format, often across multiple time periods, with an aim of uncovering the business information concealed within the data. OLAP enables business users to gain an insight into the business through interactive analysis of different views of business data that have been built up from the operational systems. This approach facilitates a more intuitive and meaningful analysis of business information and assists in identifying important business trends.

OLAP can be defined as the process of converting raw data into business information through multi-dimensional analysis. This enables analysts to identify business strengths and weaknesses, business trends and the underlying causes of these trends. It provides an insight into the business through the interactive analysis of different views of business information, that have been built up from raw operating data which reflect the business users understanding of the business. The OLAP application contains logic, which includes multi-dimensional data selection, sub-setting of data, retrieval of data via the metadata layers and calculation formulas. The OLAP application layer is accessed via a front-end query tool, which uses tables and charts to “drill down” or navigate through dimensional data or aggregated measures (see Figure 13.4).

Although OLAP applications are found in widely divergent functional areas, they all require the following key features: multidimensional views of data,

calculation-intensive capabilities and time intelligence. Figure 13.11 below illustrates the major features of an OLAP system.

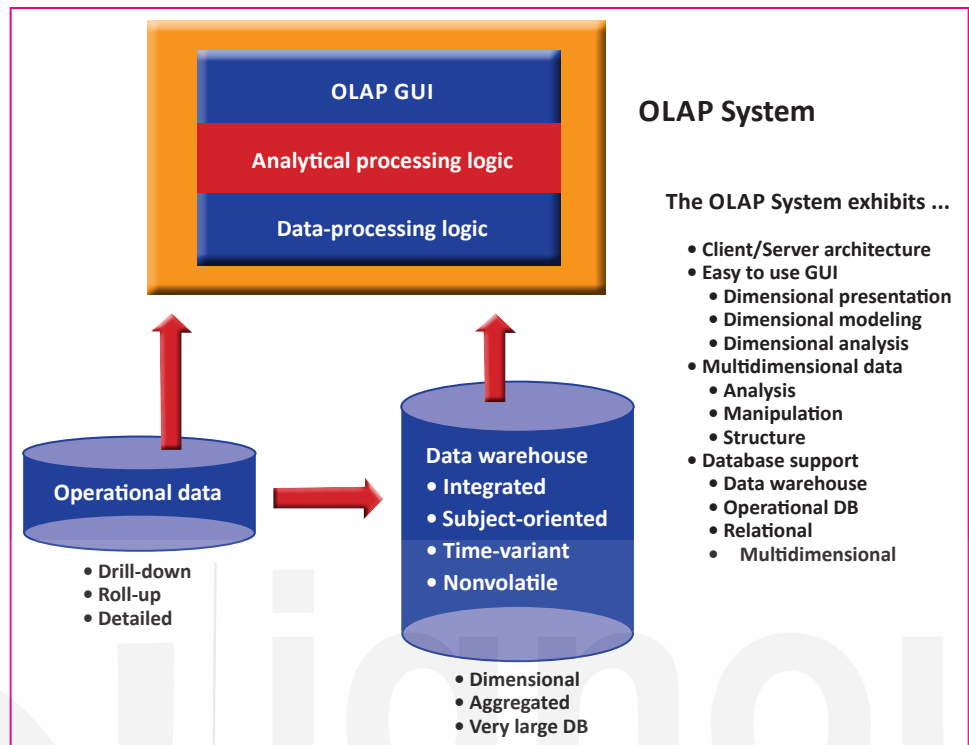


Figure 13.11: The OLAP System Architecture

The *OLAP Engine*, which is comprised of the ‘*Analytical processing logic*’ and the ‘*Data processing logic*’ layers, shown in Figure 13.11 above, provides a ‘front-end’ to the data warehouse – an interface through which a variety of general purpose *Graphical User Interface* (or *GUI*) type tools can access the data warehouse (see Figure 13.12 below). This arrangement makes the data warehouse data available for non-multi-dimensional analysis as well.

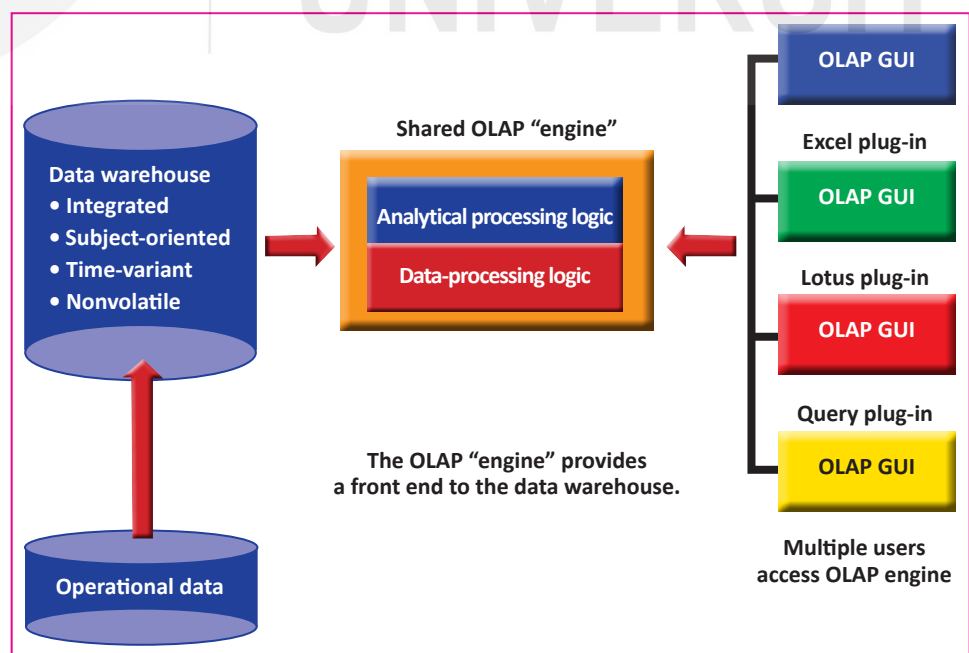


Figure 13.12: The OLAP Engine

The four major types of OLAP applications are multidimensional OLAP, hybrid OLAP, desktop OLAP and relational OLAP. Multidimensional OLAP is based on a multi-dimensional data base architecture. This stores data in a three-dimensional data cube that is already in the OLAP multidimensional format for “slicing and dicing” into analysis views. Relational OLAP products are designed to operate directly on a data warehouse built on relational databases, through a comprehensive metadata layer. Hybrid OLAP products primarily integrate specialized multidimensional data storage with relational database management technology. This allows businesses to link multi-dimensional data to the underlying source data in a relational or object oriented database. The desktop style of OLAP allows users to perform limited analysis, directly against data held within a relational or object oriented database, while avoiding many of the problems that affect hybrid and relational OLAP styles.

13.9 DATA VISUALIZATION

Business decision-makers, use data to identify opportunities, trends, and areas of concern in their respective businesses. Most data reaches the users in the form of tabular reports, which they find challenging to quickly and effectively absorb the information, spot patterns, identify aberrations, and see hidden relationships. On the other hand, the trained human eye can easily ‘spot’ patterns that the most advanced analytical tools would not be able to detect. Fortunately, though the volume of data that the users need to deal with is ever expanding, data-visualization tools have been evolving to the point that they can now transform large quantities of complex data into meaningful visual representations that incorporate the science behind human perception and cognition. As the old saying goes, a picture is often worth a thousand words, or in this case, a thousand rows of data. Data-visualization applications render large quantities of data in the form of basic charts, graphical indicators, scorecards, dashboards, advanced visualizations, animations, and virtual reality environments. By automating the fundamental aspects of data analysis, they help information users identify trends and patterns in data that are often not apparent in traditional tabular representations. Managers have found these applications very effective for viewing business activity at a glance.

Data visualization involves graphical image generation by software, where the content of the image is determined by reading digital data. The data is usually numeric, but some software can visualize concepts drawn from text documents. The software arranges geometric shapes (such as points, lines, circles and rectangles) to create an interpretation of the data it read. Attributes such as proximity, size and color express relationships between the geometric shapes. For example, Figure 13.13 below visualized Product, Location, Revenue, Number of items sold, and Distribution channel information sourced from the data warehouse of a retail chain, in a single view. Data visualization has gained adoption among business users because it supports a variety of business tasks, including decision-making, knowledge management and business performance management.

Three strong trends have shaped the direction of data visualization software for business users over the last few years:

Chart Types. Most data visualizations depend on a standard chart type, whether a rudimentary pie chart or an advanced scatter plot (like the one used in Figure 13.13 below). The list of chart types supported by software has lengthened considerably in recent years.

Level of User Interaction with Visualization. A few years ago, most visualizations were static charts for viewing only. At the cutting edge of data visualization today, dynamic chart types are themselves the user interface, where the user manipulates a visualization directly and interactively to discover new views of information online.

Size and Complexity of Data Structures Represented by Visualization. A rudimentary pie or bar chart visualizes a simple series of numeric data points. Newer advanced chart types, however, visualize thousands of data points or complex data structures, such as neural nets.

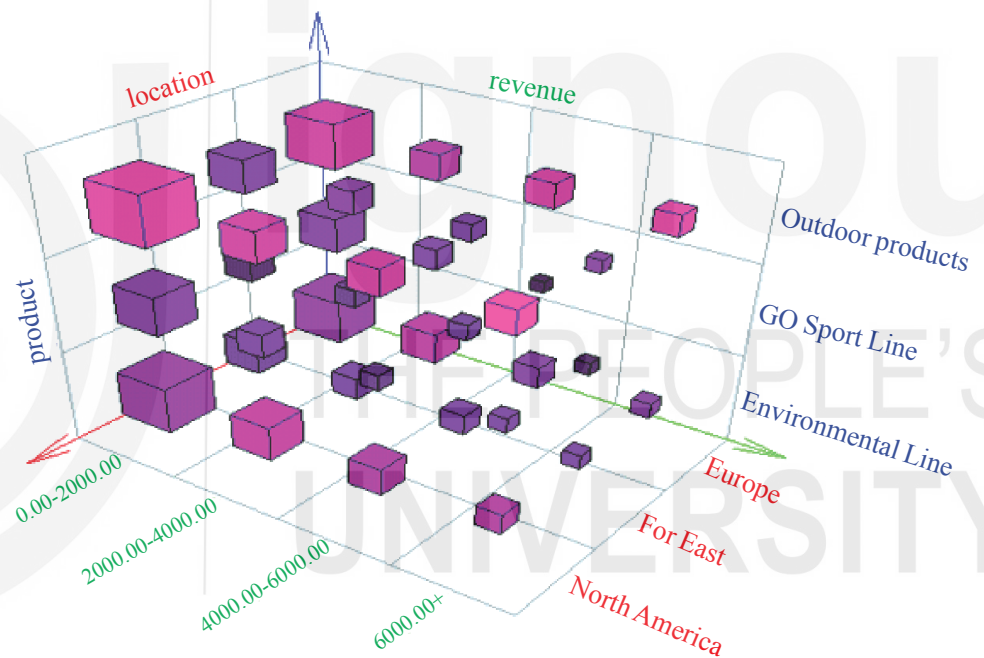


Figure 13.13: Visualization of Retail Data

Figure 13.14 depicts these trends and places in their context some of the common functionality of data visualization software. These trends also reveal a progression from forms of rudimentary data visualization (RDV) to advanced ones (ADV), as seen moving from lower-left to upper-right corners in Figure 1. Rudimentary forms of data visualization (pie and bar charts and presentation graphics) have been available in software for many years, whereas advanced forms (with interactive user interfaces, drill-down and live data connectivity) are relatively new. The dotted lines in Figure 13.14 create a life-cycle context for rudimentary and advanced forms of data visualization by identifying three life-cycle stages: maturing, evolving and emerging.

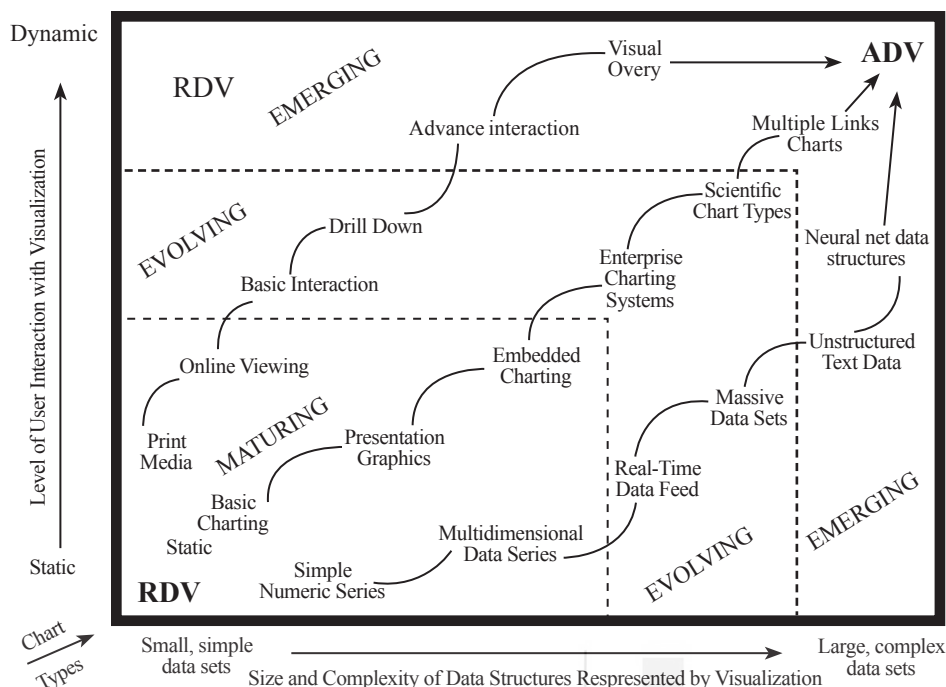


Figure 13.14: Charting – State-of-the-art

The wide array of visualization software supports many different chart types, because the needs of users vary greatly. For example, business users demand pie and bar charts, whereas scientific users need scatter plots and constellation graphs. Users looking at geospatial data need maps and other three-dimensional data representations. Digital dashboards are popular with executive business intelligence users who monitor organizational performance metrics, visualizing them as speedometers, thermometers or traffic lights.

Tools for charting and presentation graphics are devoted exclusively to visualizing data. However, data visualization capabilities are commonly embedded in a wide range of software types, including tools for reporting, OLAP, text mining and data mining as well as applications for customer relationship management and business performance management.

Business software for charting has evolved from static and superficial charts to interactive visualizations with data connectivity and drill down. These advanced capabilities are found in a new category of software the enterprise charting system (ECS). With an ECS, users can develop and deploy chart-centric analytic applications that provide data visualization specifically for business intelligence on an enterprise scale.

Data visualization has long been in use in software for the study of mathematical, statistical, geographic and spatial data. Some data visualization software for business users has borrowed chart types originally designed for scientific users, such as scatter plots and constellation graphs.

Whereas scientific users tolerate heavily technical functionality that may require knowledge of programming languages or statistics, business users need this functionality to be hidden under a friendly user interface. For data visualization to appeal to business users, it must provide out-of-the-box

value in the form of functionality for solving business problems, such as analyzing or mining customer behavior, product categories and business performance.

What is good visualization?

Effective: ease of interpretation

Accurate: sufficient for correct quantitative evaluation.

Aesthetics: must not offend viewer's senses

Adaptable: can adjust to serve multiple needs

Human perception capabilities and the fact that all displays need to be mapped to a 2-dimensional flat computer screen or sheets of paper (when printed) impose many limitations. For example, according to one researcher, inaccuracy of human perception increases in the following order when the respective techniques are used for visualization: Position along a common scale, Position along identical non-aligned scales, Length, Angle / slope, Area, Volume, Hue / saturation / intensity.

The techniques commonly used to overcome the limitations of 2-dimensional displays and size restrictions are:

- Scaling and offset to fit in range
- Derived values (e.g. residuals, logs) to emphasize changes
- Projections and other combinations to compress information
- Random jiggling to separate overlaps
- Multiple views to handle hidden relations, high dimensions
- Effective grids, keys and labels to aid understanding.

13.10 DATA MINING

According to Berry and Linoff, *Data Mining* is the exploration and analysis, by automatic or semiautomatic means, of large quantities of data in order to discover meaningful patterns and rules. This definition, justifiably, raises the question: how does data mining differ from OLAP? OLAP is undoubtedly a semiautomatic means of analyzing data, but the main difference lies in quantities of data that can be handled. There are other differences as well. Tables 5 and 6 summarize these differences.

Table 13.5: OLAP Vs Data Mining – Past Vs Future

OLAP: Report on the past	Data Mining: Predict the future
Who are our top 100 best customers for the last three years?	Which 100 customers offer the best profit potential?
Which customers defaulted on the mortgages last two years?	Which customers are likely to be bad credit risks?
What were the sales by territory last quarter compared to the targets?	What are the anticipated sales by territory and region for next year?
Which salespersons sold more than their quota during last four quarters?	Which salespersons are expected to exceed their quotas next year?

Last year, which stores exceeded the total prior year sales?	For the next two years, which stores are likely to have best performance?
Last year, which were the top five promotions that performed well?	What is the expected return for next year's promotions?
Which customers switched to other phone companies last year?	Which customers are likely to switch to the competition next year?

Table 13.6: Differences between OLAP and Data Mining

FEATURES	OLAP	DATA MINING
Motivation for information request	What is happening in the enterprise?	Predict the future based only why this is happening
Data granularity	Summary data	Detailed transaction-level data.
Number of business dimensions	Limited number of dimensions	Large number of dimensions.
Number of dimension attributes	Small number of attributes	Many dimension attributes
Sizes of datasets for the dimensions	Not large for each dimension	Usually very large for each dimension
Analysis approach	User-driven, interactive analysis	Data-driven automatic knowledge discovery
Analysis techniques	Multidimensional, drill-down, and slice-and-dice	Prepare data, launch mining tool and sit back
State of the technology	Mature and widely used	Still emerging; some parts of the technology more mature

WHY NOW?

Why is data mining being put to use in more and more businesses? Here are some basic reasons:

In today's world, an organization generates more information in a week than most people can read in a lifetime. It is humanly impossible to study, decipher, and interpret all that data to find useful patterns.

A data warehouse pools all the data after proper transformation and cleansing into well-organized data structures. Nevertheless, the sheer volume of data makes it impossible for anyone to use analysis and query tools to discern useful patterns.

In recent times, many data mining tools suitable for a wide range of applications have appeared in the market. The tools and products are now mature enough for business use.

Data mining needs substantial computing power. Parallel hardware, databases, and other powerful components are available and are becoming very affordable.

Organizations are placing enormous emphasis on building sound customer relationships, and for good reasons. Companies want to know how they can sell more to existing customers. Organizations are interested in determining which of their customers will prove to be of long-term value to them. Companies need to discover any existing natural classifications among their customers so that the each such class may be properly targeted with products and services. Data mining enables companies to find answers and discover patterns in their customer data.

Finally, competitive considerations weigh heavily on organizations to get into data mining. Perhaps competitors are already using data mining.

DATA MINING TECHNIQUES

Data mining covers a broad range of techniques. Each technique has been heavily researched in recent years, and several mature and efficient algorithms have evolved for each of them. The main techniques are: *Cluster detection*, *Decision trees*, *Memory based reasoning*, *Link analysis*, *Rule induction*, *Association rule discovery*, *Outlier detection and analysis*, *Neural networks*, *Genetic algorithms*, and *Sequential pattern discovery*. Discussion on the algorithms associated with the various techniques has been kept outside the scope of this text for two main reasons: firstly, because they are too mathematical / technical in nature, and secondly, because there are numerous, well written text books, to serve the needs of those who are specially interested in the subject. Table 13.7 below summarized the important features of some of these techniques. The model structure refers to how the technique is perceived, not how it is actually implemented. For example, a decision tree model may actually be implemented through SQL statements. In the framework, the basic process is the process performed by the particular data mining technique. For example, the decision trees perform the process of splitting at decision points. How a technique validate the model is important. In the case of neural networks, the technique does not contain a validation method to determine termination. The model calls for processing the input records through the different layers of nodes and terminate the discovery at the output node.

Table 13.7: Summary of Data Mining Techniques

Data Mining Techniques	Underlying Structure	Basic Process	Validation Method
Cluster Detection	Distance calculation in n-vector space	Grouping of values in the same neighborhood	Cross validation to verify accuracy
Decision Trees	n-ary Tree	Splits at decision points based on entropy	Cross validation
Memory-based Reasoning	Predictive structure based on distance and combination functions	Association of unknown instances with known instances	Cross validation

Link Analysis	Graphs	Discover links among variables by their values	Not applicable
Neural Networks	Forward propagation network	Weighted inputs of predictors at each node	Not applicable
Genetic Algorithms	Fitness functions	Survival of the fittest on mutation of derived values	Mostly cross validation

DATA MINING APPLICATIONS

Data mining technology encompasses a rich collection of proven techniques that cover a wide range of applications in both the commercial and noncommercial realms. In some cases, multiple techniques are used, back to back, to greater advantage. For instance, a cluster detection technique to identify clusters of customers may be followed by a predictive algorithm applied to some of the identified clusters to discover the expected behavior of the customers in those clusters.

Noncommercial use of data mining is strong and pervasive in the research area. In oil exploration and research, data mining techniques discover locations suitable for drilling based on potential mineral and oil deposits. Pattern discovery and matching techniques have military applications in assisting to identify targets. Medical research is a field ripe for data mining. The technology helps researchers with discoveries of correlations between diseases and patient characteristics. Crime investigation agencies use the technology to connect criminal profiles to crimes. In astronomy and cosmology, data mining helps predict cosmic events.

The scientific community makes use of data mining to a moderate extent, but the technology has widespread applications in the commercial arena. Most of the tools target the commercial sector. Consider the following list of a few major applications of data mining in the business area.

Customer Segmentation: This is one of the most widespread applications. Businesses use data mining to understand their customers. Cluster detection algorithms discover clusters of customers sharing the same characteristics.

Market Basket Analysis: This very useful application for the retail industry. Association rule algorithms uncover affinities between products that are bought together. Other businesses such as upscale auction houses use these algorithms to find customers to whom they can sell higher-value items.

Risk Management: Insurance companies and mortgage businesses use data mining to uncover risks associated with potential customers.

Fraud Detection: Credit card companies use data mining to discover abnormal spending patterns of customers. Such patterns can expose fraudulent use of the cards.

Delinquency Tracking: Loan companies use the technology to track customers who are likely to default on repayments.

Demand Prediction: Retail and other businesses use data mining to match demand and supply trends to forecast for specific products.

Table 13.8: Application of Data Mining Techniques

Application Area	Examples of Mining Functions	Mining Processes	Mining Techniques
Fraud Detection	Credit card frauds Internal audits Warehouse pilferage	Determination of variation from norms	Data visualization Memory-based Reasoning Outlier detection and analysis
Risk Management	Credit card upgrades Mortgage Loans Customer Retention Credit Rating	Detection and analysis of association Affinity grouping	Decision Trees Memory based Reasoning Neural networks
Market Analysis	Market basket analysis Target marketing Cross selling Customer Relationship Management	Predictive Modeling Database segmentation	Cluster Detection Decision Trees Association rules Genetic Algorithms

THE DATA MINING PROJECT

Step 1: Define Business Objectives. This step is similar to any information system project. First of all, determine whether a data mining solution is really needed. State the objectives. Are we looking to improve our direct marketing campaigns? Do we want to detect fraud in credit card usage? Are we looking for associations between products that sell together? In this step, define expectations. Express how the final results will be presented and used.

Step 2: Prepare Data. This step consists of data selection, preprocessing of data, and data transformation. Select the data to be extracted from the data warehouse. Use the business objectives to determine what data has to be selected. Include appropriate metadata about the selected data. Select the appropriate data mining technique(s) and algorithm(s). The mining algorithm has a bearing on data selection.

Unless the data is extracted from the data warehouse, when it is assumed that the data is already cleansed, pre-processing may be required to cleanse the data. Preprocessing could also involve enriching the selected data with external data. In the preprocessing sub-step, remove noisy data, that is, data blatantly out of range. Also ensure that there are no missing values.

Step 3: Perform Data Mining. Obviously, this is the crucial step. The knowledge discovery engine applies the selected algorithm to the prepared data. The output from this step is a set of relationships or patterns. However, this step and the next step of evaluation may be performed in an iterative manner. After an initial evaluation, there may be need to adjust the data and

redo this step. The duration and intensity of this step depend on the type of data mining application. If the database is being segmented not too many iterations are needed. If a predictive model is being created, the models are repeatedly set up and tested with sample data before testing with the real database.

Step 4: Evaluate Results. The aim is to discover interesting patterns or relationships that help in the understanding of customers, products, profits, and markets. In the selected data, there are potentially many patterns or relationships. In this step, all the resulting patterns are examined, and a filtering mechanism is applied so as to select only the promising patterns for presentation and use.

Step 5: Present Discoveries. Presentation of patterns / associations discovered may be in the form of visual navigation, charts, graphs, or free-form texts. Presentation also includes storing of interesting discoveries in the knowledge base for repeated use.

Step 6: Ensure Usage of Discoveries. The goal of any data mining operation is to understand the business, discern new patterns and possibilities, and also turn this understanding into actions. This step is for using the results to create actionable items in the business. The results of the discovery are disseminated so that action can be taken to improve the business.

SELECTING DATA MINING SOFTWARE TOOLS

Before we get into a detailed list of criteria for selecting data mining tools, let us make a few general but important observations about tool selection.

The tool must be able to integrate well with the data warehouse environment by accepting data from the warehouse and be compatible with the overall metadata framework.

The patterns and relationships discovered must be as accurate as possible. Discovering erratic patterns is more dangerous than not discovering any patterns at all.

In most cases, an explanation for the working of the model and how the results were produced is required. The tool must be able to explain the rules and how patterns were discovered.

Let us complete this section with a list of criteria for evaluating data mining tools. The list is by no means exhaustive, but it covers the essential points.

Data Access: The data mining tool must be able to access data sources including the data warehouse and quickly bring over the required datasets to its environment. On many occasions data from other sources may be needed to augment the data extracted from the data warehouse. The tool must be capable of reading other data sources and input formats.

Data Selection: While selecting and extracting data for mining, the tool must be able to perform its operations according to a variety of criteria. Selection abilities must include filtering out of unwanted data and deriving new data items from existing ones.

Sensitivity to Data Quality: Because of its importance, data quality is worth mentioning again. The tool must be able to recognize missing or

incomplete data and compensate for the problem. The tool must also be able to produce error reports.

Data Visualization: Data mining techniques process substantial data volumes and produce a wide range of results. Inability to display results graphically and diagrammatically diminishes the value of the tool severely.

Extensibility: The tool architecture must be able to integrate with the data warehouse administration and other functions such as data extraction and metadata management.

Performance: The tool must provide consistent performance irrespective of the amount of data to be mined, the specific algorithm applied, the number of variables specified, and the level of accuracy demanded.

Scalability: Data mining needs to work with large volumes of data to discover meaningful and useful patterns and relationships. Therefore, ensure that the tool scales up to handle huge data volumes.

Openness: This is a desirable feature. Openness refers to being able to integrate with the environment and other types of tools. The ability of the tool to connect to external applications where users could gain access to data mining algorithms from the applications, is desirable. The tool must be able to share the output with desktop tools such as graphical displays, spreadsheets, and database utilities. The feature of openness must also include availability of the tool on leading server platforms.

Suite of Algorithms: A tool that provides several different algorithms rather than one that supports only a few data mining algorithms, is more advantageous.

Multi-technique support: A data mining tool supporting more than one technique is worth consideration. The organization may not presently need a composite tool with many techniques, but a multi-technique tool opens up more possibilities. Moreover, many data mining analysts desire to cross-validate discovered patterns using several techniques.

Activity B

You are a data mining consultant in a large commercial bank that provides many financial services. The bank already has a data warehouse which it rolled out two years ago. The management wants to find the existing customers who are likely to respond to a marketing campaign offering new services.

Outline the data mining process, list the steps / phases, and indicate the activities in each phase.

In the project you are responsible for analyzing the requirements and selecting a toolset / software for data mining. Make a list of the criteria you will use for the toolset / software selection. Briefly explain why each criterion is necessary.

What could be the business benefits of the exercise? Would the effort and cost be justified?

.....
.....

BENEFITS OF DATA MINING

Without data mining, useful knowledge lying buried in the mountains of data in many organizations would never be discovered and the benefits from using the discovered patterns and relationships would not be realized. What are the types of such benefits? We have already touched upon the applications of data mining and have indicated the implied benefits.

Just to appreciate the enormous utility of data mining, the following list enumerates some real-world situations:

- In a large company manufacturing consumer goods, the shipping department regularly short-ships orders and hides the variations between the purchase orders and the freight bills. Data mining detects the criminal behavior by uncovering patterns of orders and premature inventory reductions.
- A mail order company improves direct mail promotions to prospects through more targeted campaigns.
- A supermarket chain improves earnings by rearranging the shelves based on discovery of affinities of products that sell together.
- An airlines company increases sales to business travelers by discovering traveling patterns of frequent flyers.
- A department store hikes the sales in specialty departments by anticipating sudden surges in demand.
- A national health insurance provider saves large amounts of money by detecting fraudulent claims.
- A major banking corporation with investment and financial services increases the leverage of direct marketing campaigns. Predictive modeling algorithms uncover clusters of customers with high lifetime values.
- A manufacturer of diesel engines increases sales by forecasting sales of engines based on patterns discovered from historical data of truck registrations.
- A major bank prevents loss by detecting early warning signs for attrition in its checking account business.
- A catalog sales company doubles its holiday sales from the previous year by predicting which customers would use the holiday catalog.

13.11 SUMMARY

Information for strategic decision making is of utmost importance to organizations. Due to several practical difficulties with the structure and nature of data held in operational systems in generating the required

information and availability of secondary data worldwide, the concept of data warehousing has emerged. In the data warehousing environment data collected from internal operational systems, as well as data from external sources is loaded into a subject-oriented, time-variant, integrated, and non-volatile environment. The data warehouse environment is carefully designed with time-span, granularity, aggregations, dimensionality and partitioning in mind. The initial data feed to the data warehouse as well as regular loading of data takes place in three phases: extraction, transformation and loading, and is done by Load Manager component of the data warehouse. The data in the data warehouse is organized in terms of fact data and dimension data. The star and snowflake schemas used in data warehouses support the separation of fact data and dimension data, as well as efficient query processing. Metadata or data about data is an indispensable and integral component of a data warehouse. Data collected from external sources can supplement internal data and prove extremely valuable. The concept of collecting information from the Web primarily for loading into an organization's data warehouse is known as web farming. On-line analytical processing (OLAP) technology provides multi-dimensional analysis capability to the users of a data warehouse. It also provides an interface through which other client software can access the data warehouse. Data visualization technology has made great progress in recent years. It can act as an important tool in the hands of trained users, by providing valuable support to spotting trends and / or deviations in data. Apart from OLAP and data visualization tools that are user-driven semiautomatic tools for data analysis, another class of tools, that are more automatic in their analysis and at the same time capable of accessing and using large volumes of data, have matured recently. This category of tools, known as data mining tools, have found applicability in many business area. A variety of data mining techniques and algorithms are available for data exploration and discovery of 'novel' patterns and relationships. The business benefits derived from the discoveries made by the data mining tools could be immense.

13.12 UNIT END EXERCISES

1. What are the practical difficulties of using an operational system's database to service the information needs of strategic decision making?
2. Briefly explain the importance of the Extract-Transform-Load operation, and the functions of the Load Manager component of a data warehouse.
3. Explain briefly the terms: (a) subject orientation (b) granularity (c) aggregations (d) partitioning (e) time variance.
4. In the context of a data warehouse (a) what is a concept hierarchy? (b) what do you mean by factual information? (c) what is a snowflake schema?
5. Why is it important to separate fact data from reference data in a data warehouse?

6. Why is metadata an important component of a data warehouse? What is a metadata repository and how is it used by the query manager component of a data warehouse?
7. What do you mean by 'business intelligence'? What is an OLAP Engine?
8. Explain how the progress of data visualization technology in recent years has helped data analysts and business decision makers.
9. How can you use the Web as a data source for your data warehouse? What types of information can you get from the Web? Explain briefly the steps needed to ensure that only good quality, reliable data is loaded into the data warehouse from the Web.
10. How is data mining different from OLAP? Explain briefly.

13.13 REFERENCES & SUGGESTED FURTHER READINGS

1. Adriaans, Pieter, Zantinge, Dolf, *Data Mining*, Addison Wesley Longman Ltd., Harlow, England
2. Berry, Michael J.A., Linoff, Gordon, 2001, *Mastering Data Mining*, John Wiley & Sons, Inc., New York
3. Inmon, W.H., *Using the Data Warehouse*, John Wiley & Sons, Inc., New York
4. Inmon, W.H., *Building the Data Warehouse*, John Wiley & Sons, Inc., New York
5. Inmon, W.H., Welch, J.D., Glassey, K.L., *Managing the Data Warehouse*, John Wiley & Sons, Inc., New York
6. Leon, Alexis, Leon, Mathews, *Database Management Systems*, Leon Vikas, Chennai
7. Sinha, Amitesh, *Data Warehousing*, Thomson Asia Pte. Ltd., Singapore

UNIT 14: ARTIFICIAL INTELLIGENCE APPLICATION IN DECISION MAKING

Structure

- 14.1 Introduction
- 14.2 Objectives
- 14.3 Artificial intelligence
 - 14.3.1 Brief History of AI
 - 14.1.2 Goals of AI
 - 14.1.3 Applications of AI
 - 14.3.4 Building AI
- 14.3 Expert system
- 14.4 Neural Networks
- 14.5 Genetic algorithm
- 14.6 Fuzzy Logic
- 14.7 Summary
- 14.8 Unit End Exercises
- 14.9 References & Suggested Further Reading

14.1 INTRODUCTION

The need for intelligent support systems has arisen because a need was felt to solve not-so-trivial problems by way and means of systems that are much more capable than humans. It was felt necessary to use skillfully coded real-world knowledge and rules in to a computer, which in turn simplifies the problem solving effort. Expert systems and knowledge-based systems are efforts in this direction. Evolution of techniques like artificial neural networks, fuzzy logic and genetic algorithms has really enhanced the problem-solving capacity of these systems. In this unit we will give you a brief introduction about them so that you can appreciate these concepts.

14.2 OBJECTIVES

After reading this unit you should be able to

- Explain artificial intelligence, and expert systems;
- Describe Neural Networks, Genetic Algorithms and Fuzzy Logic;
- Explain the efforts made in creation of these systems; and
- Identify and discuss their advantages in business applications.

14.3 ARTIFICIAL INTELLIGENCE

The advent of Artificial Intelligence (AI) in fact took place when the first modern computer emerged in the middle of last century. Frankly, what a computer does is that it simulates collective thinking of several humans

to process high volumes of data at a high speed and accuracy. We are witnessing a shrink in the size and cost of the computer, but there is growth in its capability. The growth in capabilities actually means that computers are becoming more and more competent of delivering more than the human abilities. These enhanced competencies are then used to deliver more power in business activities. The goal of the AI scientists had always been to develop a thinking machine that solves problems in a way that would be considered “intelligent” if done by a human. This forms the basis of our definition to artificial intelligence. Oxford dictionary defines the two terms as “artificial” means not natural or imitating nature and “intelligence” as quickness in understanding the information. The term AI was first introduced by McCarthy, “Intelligence is the computational part of the ability to achieve goals in the world”. Artificial Intelligence (AI) thus can be defined as a field that focuses on building techniques to enable computer systems to perform activities that are quick and are above human intelligence. Patterson defined artificial intelligence, as “AI is a branch of computer science concerned with the study and creation of computer systems that exhibit some form of intelligence: systems that learn new concepts and tasks, systems that can reason and draw useful conclusions about the world around us, systems that can understand a natural language or perceive and comprehend a visual scene, and systems that perform other types of feats that require human types of intelligence”. Konar defined AI as “the simulation of human intelligence on a machine, so as to make the machine efficient to identify and use the right piece of ‘knowledge’ at a given step of solving a problem”.

The need for AI has arisen because, initially, a need was felt to solve tiny and trivial problems. To counter this, it was felt necessary to use skillfully coded real-world knowledge and rules in to a computer, which in turn simplifies the problem solving effort.

Artificial intelligence is not to replace humans. It augments our abilities and makes us better at what we do. Artificial Intelligence enhances the speed, precision and effectiveness of human efforts. In financial institutions, AI techniques can be used to identify which transactions are likely to be fraudulent, adopt fast and accurate credit scoring, as well as automate manually intense data management tasks. AI is being used today in industries, hospitals & health care, education, agriculture, legal, R&D, Retailing, Banking, Household and, what not, almost everywhere.

14.3.1 Brief History of AI

As mentioned earlier, AI has roots from the time when the computer became a commercial reality. It will be interesting to have a feel as to how these developments occurred and finally took the shape in which AI is today.

Table 14.1: Brief History of AI

Sl. No.	Period	Works
1	Works before 50's	Demonstration of Turing Machine for manipulating symbols as well as numbers giving the birth of machine intelligence, emergence of cybernetics (study of communication between human & machine) as a field, new approaches to language theories, computer becomes a commercial reality, emergence of theories like information theory, Boolean algebra, switching theory and statistical decision theory

2	Works after 50's	Development of chess playing programs, works like language translation, automatic theorem proving and new programming languages. Development of FORTRAN, LISP, computational linguistics, pattern recognition & natural language processing
3	Works in the 60's	Program to play checkers, resolution as an inference method in logic; first knowledge based expert system (DENDRAL), development of a large interactive general purpose program that solves mathematical problems and simulation of complicated processes of thinking.
4	Works in mid 70's to 80's	Solving more complex problems of practical interest, expert systems that diagnose and prescribe medicines, expert systems that deals with real-life problems, uncertainty modeling, non-monotonic and spatio-temporal reasoning etc. specialized programs about narrow problem areas. Expert systems were programmed that emulate the decision-making ability of a human expert.
5	Works in 90's	IBM Deep Blue beats world chess champion, Gary Kasparov, and became the first computer to beat a world chess champion.
6	Works in 1 st decade of 21 st Century	For the first time, AI entered the home in the form of Roomba, a vacuum cleaner. AI came in the Business world till the year 2006. Companies like Facebook, Twitter, and Netflix also started using AI.
7	Works in 2nd decade of 21 st Century	In the year 2011, IBM's Watson won jeopardy, a quiz show, where it had to solve the complex questions as well as riddles. Watson had proved that it could understand natural language and can solve tricky questions quickly. In the year 2012, Google launched an Android app feature "Google now", which was able to provide information to the user as a prediction. Chatbot "Eugene Goostman" won a competition in the "Turing test." In June 2012, Goostman won a competition, promoted as the largest-ever Turing test contest, in which it successfully convinced 29% of its judges that it was human. In November 2014 Amazon Alexa, also known simply as Alexa was created. It is a virtual assistant technology developed by Amazon. It is available in English; French; German; Japanese languages. In the Year 2018, The "Project Debater" from IBM debated on complex topics with two master debaters and also performed extremely well. Google has demonstrated an AI program "Duplex" which was a virtual assistant and which had taken hairdresser appointment on call, and lady on other side didn't notice that she was talking with the machine.

Now AI has developed to a remarkable level. The concept of Deep learning, big data, and data science are now trending like a boom. Nowadays companies like Google, Facebook, IBM, and Amazon are working with AI and creating amazing devices. The future of Artificial Intelligence is inspiring and will come with high intelligence.

We can see that all these developments were the result of the desire of the human beings to solve problems that were not so trivial. To overcome the difficulties that were faced in solving these problems, they thought of putting intelligence into a computer (essentially based on the knowledge of the then society). The theme of all these efforts was around the concept of making a program intelligent, by providing it with high quality specific knowledge

about some problem area. This collaborative and collective intelligence led the machine do the tasks better than the humans. AI has taken off from there and what we are getting today is an improved build-up on the older systems. Let us see what are the goals on which AI work.

14.3.2 Goals of AI

AI works on the following goals:

1. *Emulate humans*: From the very basic definitions of AI it has become clear that AI's main goal is to identify solvable and out of the ordinary information processing problems, and solve them. In other words, AI's primary goal is to emulate human intelligence and not only emulate but do things in a better way.
2. *Solve problems that require intelligence*: Be it a program that plays chess or a system that has all the intelligence of a dictionary, the primary aim of the AI has always remained to study and solve problem that are knowledge intensive and require a high degree of intelligence.
3. *Develop expert systems on Real-life problems*: AI has real-life problems in the world as its pivot. AI's goal is to develop experts systems that simulate the real-world problems and solve them automatically.
4. *Enhance interaction*: AI works for enhancing human-human, human-computer and computer-computer interaction. Computers have now accepting more and more varied use-interfaces for input and output processes like voice, vision and virtual reality etc.
5. Creating some system which can exhibit intelligent behavior, learn new things by itself, demonstrate, explain, and can advise to its user.

14.3.3 Applications of AI

AI is employed in virtually every branch of science today. But before going into the details of applications where AI plays a significant role, we will go back to how computers were developed and what were their capabilities at different point of times. Following table gives these details:

Table 14.2: Generations of computers based on Konar, Amit (1999)

S. No	Generation	Description	Capability
1	First Generation	Inspired by Babbage's analytical engine, these computers were made of thermo-ionic valves.	Could perform "number crunching operations"
2	Second Generation	These computers were made of transistors and were smaller in size as compared to their predecessors.	Mainly used for commercial data processing and payroll creation.
3	Third Generation	These computers were made of Integrated circuits (IC). Many electro-mechanical robots were based on these.	Could perform massive computations in real-time

4	Fourth Generation	These computers came up with high-speed VLSI engines. Many electronic robots were based on these.	Operations became more fast.
5	Fifth Generation	Emergred in the period 1981-90 in Japan with AI embedded in them. In this generation machines are being created with more and more built-in intelligence. This has generated new fields and concepts like Machine Learning, Deep Learning, Big Data Analytics, Natural Language Processing, Internet of Things etc.	Could process “intelligence”. Could process natural languages, Play games, recognize images and prove theorems. Could support in diagnostics, work as technical/virtual personal assistant. Capable of learning and self-organization.

One can see that there is a growth in the capabilities of the computers as they move from one generation to the other. In fact, this is addition of more and more intelligence in the systems that are doing wonderful works presently and are poised to do even better in the future.

You can find a lot of varied applications of AI. It has been applied in game playing, automated grammar checking, speech recognition, vision systems, and expert systems like weather forecasting, diagnostic systems, financial decision making and classification systems, computer simulated natural processes, mathematical theorem proving, natural language understanding, intelligent control and scheduling & planning. Let us consider some of the examples of these applications:

1. AI in Game Playing: Chess is one such game for which many programs are created using artificial intelligence. Chess has about 10^{120} states in a typical game. Many of such programs have done remarkably well like getting a grand-master status or getting a high ELO rating etc. Most of these programs were capable of thinking move up to five to six levels ahead. Feeding all possible moves into the computer and making it capable to search the best possible move according to the situation was the strategy of these programs. You make like to recall that Deep Blue Chess program beat world champion Gary Kasparov.
2. AI in Automated Grammar Checking: You must have used a word-processor and used its spelling and grammar-checking feature. While spelling checker module works well, there is lot that has to done for the grammar-checking module. The features available in current software are far less than satisfactory. These software fumbles at using proper pronouns, gender or giving proper meaning etc. For an efficient program of this kind, it has to be given “intelligence” about knowledge from the real world.

3. AI in Speech Recognition: You must have seen the software that types the words as one speaks. These are called phonetic typewriters. Speech understanding has found good commercial value in the recent times e.g. speech recognition systems are used in reservation through telephone. Dragon Naturally speak, are a few software that needs a mention.
4. AI in Natural Language Understanding: To understand a natural language like English one has to understand the syntactic (analysis of grammar, compilation) and semantic (meaning of sentences from association of words) interpretation of words. Robots have been designed that understands few instructions, but days are not far when a robot will fully understand a speech in the natural language. There have been examples where full document have been translated in many different languages. Site's <http://babelfish.altavista.com/> translation of web pages, is also based on natural language processing.
5. AI in Vision Systems: Any image is a two-dimensional array of pixels. In vision systems, image is processed for noise and partitioned into objects of interest. These objects are recognized and finally the image is interpreted. Many banks uses face recognition programs. There are systems like "intelligent character readers" which recognizes handwriting. Computer vision has many other applications too such as baggage inspection, manufacturing inspection etc.
6. AI in Expert Systems: An expert system consists of a knowledge base, database and an inference engine for interpreting the database using the knowledge embedded in the knowledge base. There is sound reasoning process that has to build in order to create an expert system. AI techniques have played significant role in creation of expert systems for weather forecasting; diagnostic systems such as medical diagnosis systems, pathology diagnosis system & customer assistance systems etc.; financial decision making systems like fraud detection systems used in credit card companies & systems that expedite financial transactions and classification systems like financial decision making systems & NASA's galaxy-classification system.
7. AI in Computer Simulated Natural Processes: Applications of AI are not limited to human and natural processes, it is extended to other processes too. A model based on AI tools predicts the walk of a cockroach and its basic behaviors. The model correctly predicted many of behavioral processes pertaining to cockroach.
8. AI in Mathematical Theorem Proving: There had been many attempts of proving mathematical theorems with the help of AI tools. Inference methods are used to prove new theorems.
9. AI in Intelligent Control: firstly an expert controller sets a set of rules. The rule whose premise matches the dynamic plant parameter response is selected and implemented. Fuzzy logic is also used in such cases in many industrial plants e.g. the power control in a nuclear reactor.

There are application using fuzzy logic and artificial neural networks for plant estimation i.e. for designing a process estimator.

10. **AI in Scheduling & Planning:** You must be aware of the importance of planning a time schedule of a set of events to improve the efficiency. Artificial neural nets and genetic algorithms have been employed to solve these problems. Typical examples of these applications are class-routine scheduling problem, Automatic scheduling for manufacturing and rerouting scheduler for airlines etc. Operation Desert Storm used these systems for planning logistics of people and supplies
11. **AI in Astronomy:** Artificial Intelligence can be very useful to solve complex universe problems. AI technology can be helpful for understanding the universe such as how it works, origin, etc.
12. **AI in Healthcare:** Healthcare Industries are applying AI to make a better and faster diagnosis than humans. AI can help doctors with diagnoses and can inform when patients are worsening so that medical help can reach to the patient before hospitalization.
13. **AI in Finance:** AI and finance industries are the best matches for each other. The finance industry is implementing automation, chatbot, adaptive intelligence, algorithm trading, and machine learning into financial processes.
14. **AI in Social Media:** Social Media sites such as Facebook, Twitter, and Snapchat contain billions of user profiles, which need to be stored and managed in a very efficient way. AI can organize and manage massive amounts of data. AI can analyze lots of data to identify the latest trends, hashtag, and requirement of different users.
15. **AI in Automotive Industry :**Some Automotive industries are using AI to provide virtual assistant to their user for better performance. Such as Tesla has introduced TeslaBot, an intelligent virtual assistant. Various Industries are currently working for developing self-driven cars which can make your journey more safe and secure.
16. **AI in Data Security:** The security of data is crucial for every company and cyber-attacks are growing very rapidly in the digital world. AI can be used to make your data more safe and secure. Some examples such as AEG bot, AI2 Platform, are used to determine software bug and cyber-attacks in a better way.
17. **AI in Robotics:** Artificial Intelligence has a remarkable role in Robotics. Usually, general robots are programmed such that they can perform some repetitive task, but with the help of AI, we can create intelligent robots which can perform tasks with their own experiences without pre-programmed. Humanoid Robots are best examples for AI in robotics, recently the intelligent Humanoid robot named as Erica and Sophia has been developed which can talk and behave like humans.
18. **AI in Agriculture:** Agriculture is an area which requires various resources, labor, money, and time for best result. Now a day's

agriculture is becoming digital, and AI is emerging in this field. Agriculture is applying AI as agriculture robotics, solid and crop monitoring, predictive analysis. AI in agriculture can be very helpful for farmers.

19. AI in E-commerce: AI is providing a competitive edge to the e-commerce industry, and it is becoming more demanding in the e-commerce business. AI is helping shoppers to discover associated products with recommended size, color, or even brand.

14.3.4 Building AI

While building artificial intelligence in an organization one has to think about many factors such as observation, reasoning and action. Take the case of a human being; it has sensory functions attached to various parts of its body with which it observe the things in the real world. You might have seen different software for medical diagnosis, e.g. homeopathic software contains a set of symptoms and test results that have been obtained and inputted to the system manually. Whenever a patient inputs its symptoms then the diagnosis system matches it with the recorded ones to give a diagnosis. The second step in this is the reasoning part. This is dependent on the inference, decision-making and classification from what is observed in the first step. There are many tools available for this step. These tools could be artificial neural network, logical deduction system, heuristic searching a problem space, Bayes network inference, genetic algorithms etc. The third and the final step in the building process is the action. We will go back to the example of the human beings. When the observation is done and reasoning is made it is the time for action e.g. if you touch a hot thing, you develop a reasoning that if you touch that, you will be hurt. The action is that you would not touch a hot thing. In real world you may come across different kind of problems. There would be problems that could be not solved by mathematical or logical algorithms. Such problems can only be solved through intuitive algorithms. These problems are the problems that are deemed in to the AI domain. The key to AI approach is intelligent search and matching. If you were given a problem you would be doing the following tasks:

1. Knowledge Representation and Reasoning: You will first try to represent the facts in some way. Logic is built to answer questions like “what” & “how”. One tries to collect all facts whether missing, implicit, explicit or probabilistic to gather knowledge to solve a problem. You build rules to infer from the knowledge thus collected. In case you have many ways to infer something then you would like to go for the path that is al the smallest distance from the goal.
2. Heuristic search: One has to develop rules of thumb to solve the constraints in the problem space. As we have earlier mentioned that there are about 10^{120} states in a Chess game. As we move from one state to another state we search for the best possible solution at each state. The better the solution found at each state, the fewer steps it will require to reach to the next state.

3. Inference: While searching you would come across many rules. In many cases you will be able to infer some new rules from some facts. You can build reasoning to seize a fact from other facts. The reasoning could be non-monotonic reasoning or reasoning under uncertainty.
4. Learning: Learning could be parametric learning, inductive learning and analogy-based learning. In parametric learning the adaptive learning process adjust the parameters of the subject that is learning. In inductive learning the learner makes generalizations from the examples and in analogy-based learning the learner makes comparisons form another similar problem or take the benefit of resemblance of this situation with others.
5. Planning: While a reasoning problem works for satisfying a particular condition from a given set of data and knowledge. A planning problem works for a bigger goal. It deals with the determination of a methodology for achieving a goal form the initial stages. It moves from general facts to basic facts, from basic facts to particular facts about a situation. Planning stage create an approach for achieving goals in terms of a sequence of primitive steps or actions.

14.4 EXPERT SYSTEMS

Amongst the most noteworthy developments in the field of Artificial Intelligence (AI) is the advent of “expert” or “knowledge based” systems. Joint efforts by human experts yielded systems that can diagnose diseases, fly planes, drive vehicles and configure computer systems at performance levels that can exceed the best human expertise. Question thus arises that what are expert systems? To put it most simply: Expert systems are computer programs that use knowledge to solve problems competently and successfully. They are similar to human experts in the sense that they also use logic and heuristics to solve problems, they also make errors and they also learn from their errors. This expertise is easier to store, retrieve, transfer and is cost-effective and permanent. Johnson (1983) described the term ‘expert’ in the most accurate manner as, “ An expert is a person who, because of training and experience, is able to do things the rest of s cannot; experts are not only proficient but also smooth and efficient in actions they take. Experts knows a great many things and have tricks and caveats for applying what they know to problems and tasks; they are also good at plowing through irrelevant information in order to get at basic issues, and they are good at recognizing the problems they face as instances of types with which they are familiar. Underlying the behavior of experts is the body of operative knowledge, we have termed expertise...” Thus we can now define expert systems. Patterson (1990) described expert systems as, “ An expert system is a set of programs that manipulate encoded knowledge to solve problems in a specialized domain that normally requires human expertise. An expert system’s knowledge is obtained from expert sources and coded in a form suitable for the system to use in its inference or reasoning process.” You might recall the earlier description; an expert system consists of a knowledge base, database and an inference engine for interpreting the database using

the knowledge embedded in the knowledge base. There is sound reasoning process that has to build in order to create an expert system.

Brief History of Expert Systems

Expert systems emerged as a consequence of the developments in the artificial intelligence field in early 70's at a few leading US universities like Stanford. They started as problem solvers using specialized domain knowledge. We will discuss a few early successful systems. This is summarized in the table given below:

Table 14.3: Brief History of Expert Systems
(Based on inputs from Patterson, 1990 and Waterman, 1986)

S. No.	Expert System	Year	Developer	Functions
1	DENDRAL / Meta-DENDRAL	Late 60's	Stanford University	Determines the structure of chemical compounds using constituent elements and mass spectrometry data. Later adapted inductive learning form
2	MYCIN/ THEIRESIUS/ GUIDON/ EMYSIN	Mid 70's	Stanford University	Diagnoses infectious blood diseases and determine therapies. Started with 200 rules to build over 600 rules by early 80's
3	PROSPECTOR	1974-1983	Stanford Research Institute	Assists geologists in the discovery of mineral deposits. First computer system to assist geologists.
4	XCON	Late 70's	Digital Equipment Corporation & Carnegie-Mellon University	Select and configure components of complex computer systems. Expert system work in computer systems is typified by XCON.
5	ACE	Early 80's	Bell Laboratories	Equipment fault diagnosis and integrated circuit design. AT&T uses it for identify trouble spots in telephone networks
6	HASP/ SIAP	Early 70's	Stanford University & Systems Control Technology	Identifies ship types by interpreting data from hydrophone arrays that monitors regions of the ocean

Some more examples of expert systems are as below:

PXDES: Pneumoconiosis X-Ray Diagnosis Expert System (PXDES) is an expert system which is used to diagnose Lung diseases. It is an expert system that is used to determine the type and level of lung cancer. To determine the

disease, it takes a picture from the upper body, which looks like the shadow. This shadow identifies the type and degree of harm.

CaDeT: Cancer Detection (CaDeT) is a computer based diagnostic and decision support system that can detect cancer at early stages.

These were some of the path setting work in the expert systems. These works provided the base for other systems to build on them. Today one can see a plethora of applications of expert systems in virtually all areas like agriculture, chemistry, computer systems, electronics, engineering, geology, information management, law, manufacturing, mathematics, medicine, meteorology, military science, physics, process control and space technology etc.

Working Principles of Expert Systems

Expert System is a result of the interaction between the system builder (knowledge engineer) and many domain experts.

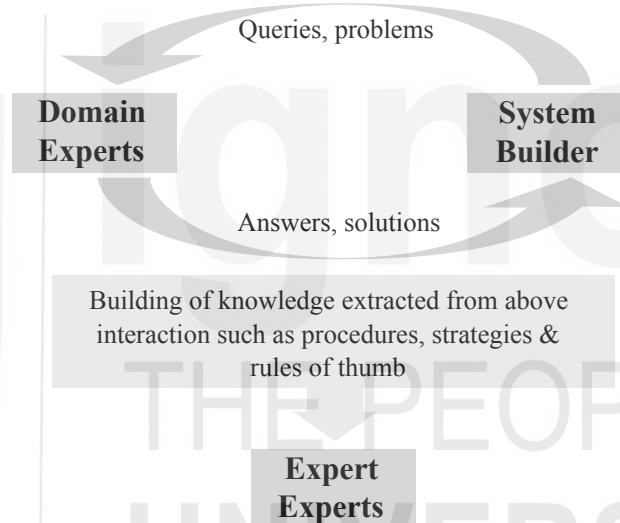


Figure 14.1: Building knowledge into an expert system (Based on Waterman, 1999)

Expert systems are computer systems that are based on knowledge rather than the data. They accumulate this knowledge at the time of system building. Knowledge is programmed and kept in such a manner so that it can be browsed and appended from time to time. Expert systems possess a very high level of expertise in the area for which they are made for. The best thing about expert systems is that they grow over time and but for the initial expenditure incurred in building them, they work in a cost effective manner. Expert systems have predictive modeling power i.e. they are capable of describing the effects of new situation on the data and the solution. Expert system uses symbolic representations for knowledge (rules, networks or frames). This compilation often becomes a quick reference for best strategies, methods and consensus decisions. This becomes a permanent knowledge base. Thus, one can say that expert systems have a permanent memory. Expert systems gives access to the user to understand it's reasoning and can be used to provide training. This is possible because of its knowledge base; it can provide trainees with experiences and strategies from which to learn.

Building Expert System

The process of building an expert system is also called knowledge engineering. As we have discussed earlier that there is an interaction between the domain expert and the expert system builder which generates the knowledge into the expert system. The latter is also called “knowledge engineer”. There is further to it. The remaining players in the building of the expert system are the expert system building tools, the user and the expert system itself. Let us have a quick tour on their meaning. Expert system is the collection of programs that solves problems in a particular area; domain expert is a person who is knowledgeable about that particular problem area and has solutions to those problems; knowledge engineer is a computer science expert who translates the domain expert’s knowledge into the one, which is understandable by the computer; expert system building tool is the programming language that is used by the knowledge engineer to build the expert system. And finally the user is the one, which will use the expert system.

Waterman (1999) has highlighted the relationship between the two in the Figure 19.2.

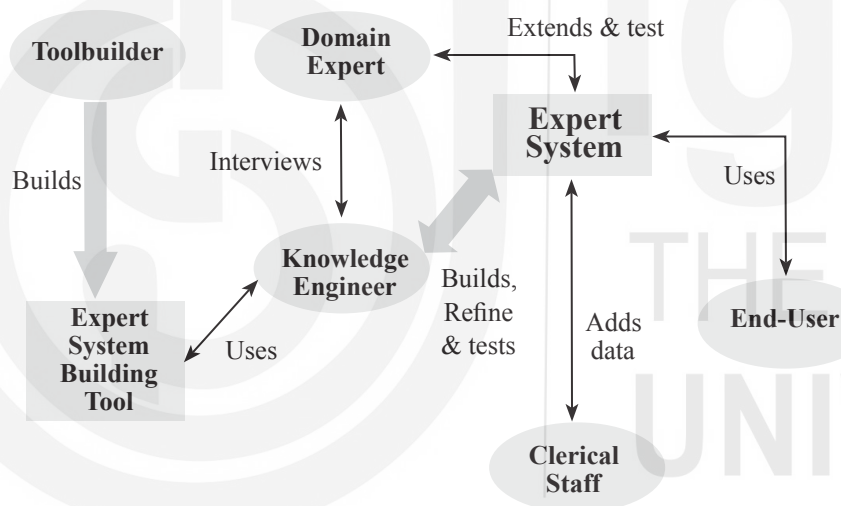


Figure14.2: The players in the expert system game (Figure adapted from Waterman, 1999)

Waterman further stated, “ The figure suggests that the user may be a tool builder debugging the expert-system-building tool includes both the language, a knowledge engineer refining the existing knowledge in the system, a domain expert adding new knowledge to the system, an end-user relying on the system for advice, or a member of the clerical staff adding data to the system”.

Knowledge representation in Expert Systems

The knowledge can be represented in expert systems in three ways. These ways are rules, frames and semantic nets. A rule-based system consists of a rule-base (permanent data); an inference engine (process); and a workspace or working memory (temporary data). Knowledge is stored as rules, which are of the form

IF some condition THEN some action

For example

IF it is hot THEN turn the AC on

IF the temperature is more than 35° THEN it is hot

IF the weather is not so hot but it is humid THEN also turn the AC on.

When the IF condition is satisfied by the facts, the action specified by the THEN is performed. These rules may modify the set of facts in the knowledge base. These new facts can be used to form matches with the IF portion of the rules. This matching of rule IF portions to the facts can produce “inference chains”. The inference chains displays how a conclusion is reached in an expert system. Rules can be used in two ways: - by forward chaining and backward chaining.

Forward Chaining: Suppose you have a database which has some information stored in it and you want to generate some new rules and test them with respect to the database. Then by the forward chaining method you start with a fact that is present in the database. You infer some rule that comes due to a consequence of a given fact. Next is that you add that inference as fact in the database. In other words you have appended the database. Once a fact is added, it interacts with other facts and infers some new fact that again gets added into the database. This process is continued till there is no further inference. The process is “forward” because it uses the facts on the left side to derive information on the right hand side.

Backward Chaining: Suppose you want to test whether a fact exists in the database. The process first checks the database for that fact. If the fact that you want to establish is not there, then the process tries to find other facts that conclude that fact. It tries to establish those facts first, which conclude the fact you are interested in. The process continues in the same manner till sufficient evidence is found for existence/non-existence of the fact being tested.

The essential difference between forward chaining and backward chaining is in the way the facts are searched in the database. In some situations, backward chaining proves to be cost-effective. If your intentions were to infer a particular fact then by employing forward chaining you would waste a lot of time and efforts. This is because of the fact that experts systems are generally based on hundreds and thousands of facts. On the other hands if you were looking for some new information from the database then by forward chaining you would be able to find a lot of rules that will be executed as a consequence of the existing facts. Forward chaining can derive a large number of inference chains and situations. These facts and rules could be further used to append the database.

Activity A

1. You must have used a word-processor and used its spelling and grammar-checking feature. What additional features would you intend

to have for the grammar-checking module? Mention the limitations you have observed in your word-processors module.

.....

.....

.....

.....

.....

.....

2. Mention five activities that you encounter in your day-to-day life that could be not solved by mathematical or logical algorithms. Take one such example and solve it (up to 2-3 stages only) by intuitive approach as done in AI systems.

.....

.....

.....

.....

.....

14.5 NEURAL NETWORKS

Neural networks are a great aid to computing. They involve mathematical structures that are capable of learning. A neural network is a set of connected input/output units where each connection has a weight associated with it. Neural networks are primarily used for predictions. Neural networks use past data and fit a model on it to predict and classify. Neural networks, as the name suggests, resemble the namesake found in human beings. Neural networks begin with an input layer, which is given some weight before it connects to the output unit. The output unit processes the values of input variable and weights with a combination function. There are many hidden layers between the input and output layer. A neural network is trained by assigning weights on the inputs of each of the units so that the network predicts the variable under question in the best possible manner. According to Barry & Linoff (2001), "Neural Networks are a good choice for most classification and prediction tasks when the results of the model are more important than understanding how the model works. Neural Networks actually represent complex mathematical equations, with lots of summations, exponential functions and any parameters. These equations describe the neural network, but are quite opaque for human eyes... Neural networks do not work well when there are many hundreds or thousands of input features... Neural networks work well with decision trees... decision trees are good at choosing the most important variables- and these can be used for training a network."

A major difference between neural networks and the statistical modeling is that the former is based on the rules it explores from the data while the latter is based on previously specified functional form.

As our nervous system is made up of neurons that are connected with a nerve fiber, neural nets also have nodes that are connected. As a neuron receives signals from its neighboring neurons, processes them and move those signals forward, the neural net also do the same functions.

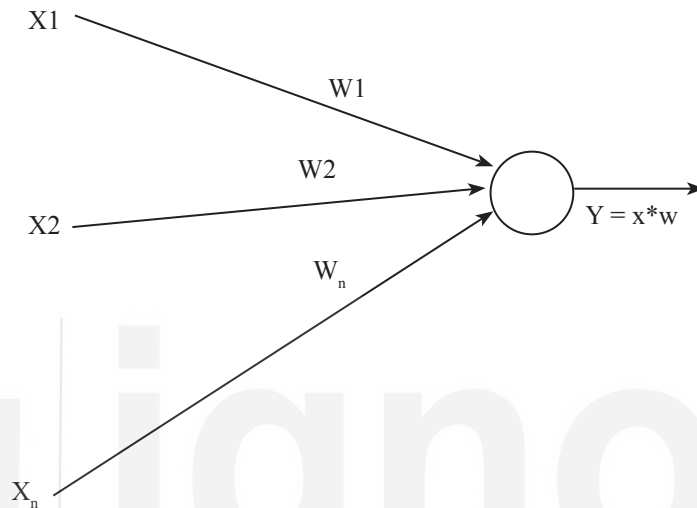


Figure 14.3: A Single Node Neural Network

Thus artificial neural networks are electrical analogues of the biological neural nets. The most common application of neural network is in the field of machine learning. One can find many applications of neural networks in control, automation, robotics, computer vision, scheduling, knowledge acquisition and planning. They are useful in many applications because they derive meaning from complicated or imprecise data. These situations are otherwise complex and undetectable by humans and other computer techniques. Neural Networks are good at identifying patterns or trends in data and hence they are well suited for prediction and forecasting needs.

A neural network can have multiple layers. Suppose, we have a training data and we measure attributes for training data. These measured attributes form the input layer. Weights are given to the input layer according to their importance in the problem. The output from the input layer is assigned some weights and this feed to the second layer, known as hidden layer. Similarly the weighted outputs of this hidden layers can be fed to the next hidden layer and so on. Finally these outputs go into the output layer, which emits network's prediction. This process is explained in the figure given below. Figure shows how a training sample $x_1, x_2, \dots, x_i$ is fed to the input layer. Weighted connections exist between each layer denoted by w_{ij} .

If there is a single hidden layer then the network is called a two-layer neural network. Similarly a network containing two hidden layers is called a three-layer neural network.

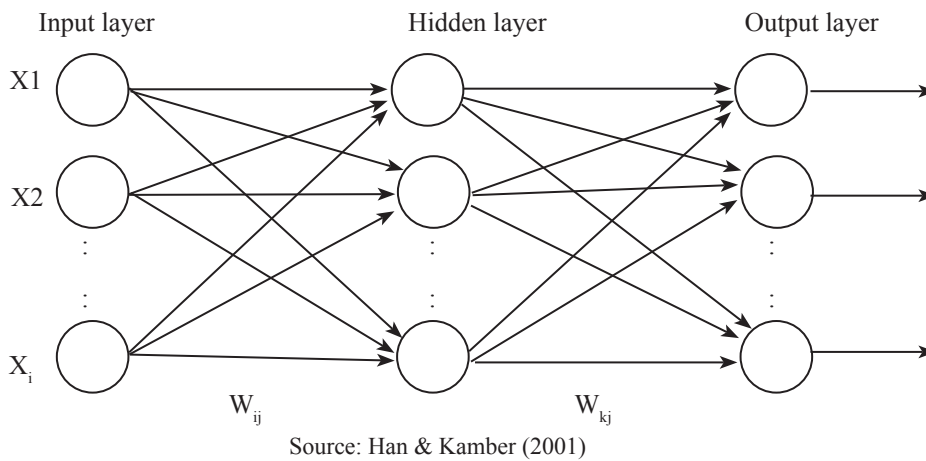


Figure 14.4: A Multilayer Neural Network Source

Business applications of Neural Networks

There are immense applications of Neural Networks in business. It is just not possible to mention all of them. However an attempt has been made to give you a glimpse of the purposes for which neural networks are being applied. Following are some of the applications of the artificial neural networks Pujari (2001), Konar (2000):

1. Used in investment analysis to predict the movement of stocks, currencies etc., from previous data. They have become worthy successors of linear models.
2. Used in monitoring the state of aircraft engines. Parameters like vibration levels and sound are used to predict or forecast engine problems.
3. Used to improve marketing mailshots. They are employed to find a predictive mapping from the known data about the clients and their response. This mapping is then used to direct further mailshots.
4. A simple neural network based diagnostic expert system, which diagnoses and recommends treatments for acute sarcophagal disease.
5. In the development of a system for criminal investigation for voice classification scheme. A multi layered feed forward neural net was trained. Speech features of the suspect are supplied to the neural net, which then processes the output signals.
6. In path planning of the mobile robots. Mobile robots works on sensors that help them recognize the world as the humans do. Neural Network based navigational model has been applied for online navigation of such robots. They help the robot to receive sensory information, generate control commands for motion and direction and finally set the schedule of motions.
7. For acquisition of knowledge in an expert system. Fuzzy neural networks are applied for automated estimation of certainty factors of knowledge from proven and historical databases of a typical reasoning system.

8. In recognition of human faces from their facial images. Self-organizing neural nets are applied for this task. Weights are given to each multi-dimensional point on the 2-d plane and then it is mapped to a single point. This process is repeated for all the 360 images (corresponding multi-dimensional points.)
9. In cognitive learning for applications in psychological modeling. Fuzzy neural networks are used to mimic the activities of “long term memory” of the biological cognitive system.

These applications are never-ending. As said earlier, neural networks are useful in many applications because they derive meaning from complicated or imprecise data. These situations are otherwise complex and undetectable by humans and other computer techniques. We now move on to another important concept i.e. fuzzy logic and its business applications.

14.6 FUZZY LOGIC

Traditional logic has many limitations when one deals with uncertain and incomplete knowledge. Unfortunately knowledge in the real world is never certain and complete. Fuzzy logic came into picture as a concept that extends the expressive power of traditional logics. You would recall that in standard set theory, an object is either a member of a set or it is not. There is no middle way. There is no partial containment. The characteristic function of a traditional set assumes values 1 and 0, on the basis of membership. If we generalize this concept and allow the characteristic function to assume all values between 0 and 1, then we talk about the fuzzy sets and the value assumed by the characteristic function denotes the degree or level of membership.

Prof. Zadeh of University of California, Berkley in 1965, first introduced fuzzy sets. The main objective was to define the fuzziness and vagueness. Fuzzy set theory forms the basis of Fuzzy logic. It is a relatively new discipline that has found applications in automated reasoning of expert systems. Fuzzy models require defining of memberships functions. These functions can often be defined on the basis of intuitive knowledge. Because of their simplicity to use and cost effectiveness, fuzzy logic and fuzzy computation are good prospects for the next generation expert systems. The universe of fuzzy set can take any value in the real interval 0 and 1. Let us take an example to explain the concept better. Suppose we want to define the membership of batsmen of national level in a country. We may define the membership characteristic function u as follows.

$$u_A(x) = \begin{cases} 1 + \left[\left(\frac{50-x}{10} \right)^2 \right]^{-1} & \text{for } 0 \leq x \leq 50 \\ 1 & \text{for } x > 50 \end{cases}$$

here x is the national average of the batsman. The interpretation is if a batsman has an average of more than 50 runs he is strong member of the fuzzy set. An average of 40 means his membership is partial with $u_A(x)=0.5$,

similarly an average of 30 would get $u_A(x)=0.2$ and so on. The characteristic function for fuzzy sets provides a direct linkage to fuzzy logic.

Business applications of Fuzzy Logic

1. Fuzzy matching is used in matching problems that is an essential operation in speech recognition, automated reasoning and expert systems. Computing fuzzy distance between two objects forms the basis of such matching.
2. Fuzzy relations are used in modeling a physical system, whose variations of output parameters with input parameters are known.
3. Fuzzy logic has been applied in management of ambiguity of the data by analyzing the knowledge about predecessors to determine the truth-value of a proposition.
4. Fuzzy reasoning is used for identifying criminals from imprecise and inconsistent word description of criminal history.
5. Fuzzy logic is used in medical diagnostic systems in identifying diseases on the basis of symptoms by calculating fuzzy beliefs.

There are many other uses of Fuzzy logic like its applications in specialized belief networks called Fuzzy Petri Nets (FPN) but are beyond the scope of this introductory unit.

14.7 GENETIC ALGORITHMS

Genetic algorithms are stochastic algorithms. They were first proposed by Holland in 1975. They are based on the natural process of biological evolution. These methods follow the “Survival of the fittest” principle of Darwin. They are called “genetic” because they work on the same principles as genetics: The members, which adapt well to the environment, are selected for reproduction and produce offspring. Poor performers die without offspring.

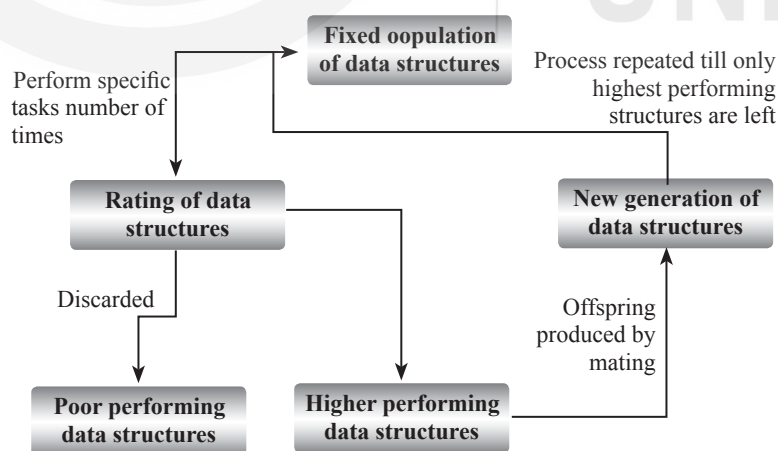


Figure 19.5: Process of Genetic Algorithms

The figure above can be summarized that genetic algorithms let the data structures face the environment and carry forward with only the fittest data structures. There are a few concepts that need to be highlighted.

Mating: Mating is done by randomly selecting a bit position in the eight bit string and concatenating the head of one parent to the tail of the other parent to produce the offspring e.g. the fifth bit crossover point of parents xxxxxxxx & yyyyyyyy will produce offspring xxxxyyyy & yyyyyxxx.

Inversion: Inversion is done on a single string. The inversion concatenates the tail of the string to the head of the same string e.g. sixth bit inversion of string abcdefgh will be ghabcdef.

Mutation: In mutation bits at one or more randomly selected positions are altered. It is done to ensure that all locations of the rule space are reachable so as to avoid local minima.

Business applications of Genetic Algorithms

1. Genetic algorithms have been used in intelligent search problems. These problems include games, routing problems in VLSI circuits and navigational planning of robots.
2. Genetic algorithms have been used in many optimization problems like the traveling salesman problem and the job shop scheduling problems etc.
3. Genetic algorithms and artificial neural networks together have many applications in machine learning such as adaptive control etc.
4. Genetic algorithms are used in data mining for the tasks of hypothesis testing and refinement, optimal decision tree induction and event prediction etc.
5. Genetic algorithms are also used in classification problems and for evaluating the fitness of other algorithms.

Activity B

1. Identify and describe two good applications areas for artificial neural networks and fuzzy logic within a university environment.
.....
.....
2. Suppose you want to define the membership of young people. Think of a membership characteristic function μ and calculate its value for different ages. You can take a person as young if he is less than 30 years of age. Also, plot the values of the membership function with respect to age.
.....
.....

14.8 SUMMARY

To overcome the difficulties that were faced in solving many real-life problems, intelligence was put into a computer (essentially based on the knowledge of the then society). The theme of all these efforts was around the concept of making a program intelligent, by providing it with high quality specific knowledge about some problem area. This collaborative and collective intelligence led the machine do the tasks better than the humans.

AI has taken off from there and what we are getting today is an improved build-up on the older system but with more and more applications because of constant developments in the tools required for deploying such AI based systems. With machine Learning, Deep Learning, NLPs we have come across many such uses of AI in almost every field – from R&D, Healthcare, Industry to Agriculture, Retailing, and eCommerce. Expert systems emerged as a consequence of the developments in the artificial intelligence field. They started as problem solvers using specialized domain knowledge. As the time passed by these systems kept building on themselves and what we are getting today is much better versions of their predecessors. More and more algorithms and methods were developed to solve problems more efficiently. Some of these include artificial neural networks, fuzzy logic and genetic algorithms. They have immense applications in the field of AI and expert systems. Genetic algorithms are used with neural networks and fuzzy logic for solving more complex problems. They are jointly used and hence are often referred as “soft-computing”. This unit has tried to provide you with an introduction of these developments.

14.9 UNIT END EXERCISES

1. What do you understand by Artificial Intelligence? What are the factors that highlighted the need of developing AI? What are the goals of AI?
2. Discuss the application of AI with respect to the developments in the field of computers. Also discuss the major works that typically describes the earlier work done in AI.
3. If you have to build artificial intelligence in your organization, what factors you would think of and take into consideration. Mention those factors in a stepwise manner.
4. Define Artificial Intelligence and mention some of its current applications.
5. What are expert systems? Mention the working principles of expert systems. Also discuss how knowledge can be represented in expert systems.
6. What are artificial neural networks? Discuss their business applications.
7. Discuss the salient features of the genetic algorithms. How do they differ from the artificial neural networks? Explain mutation and mating with the help of examples.
8. What are the factors that differentiate traditional logic and fuzzy logic? Explain the concept of membership characteristic function and also discuss the Business applications of Fuzzy Logic

14.10 REFERENCES & SUGGESTED FURTHER READING

1. Berry Michael. J.A. & Linoff Gordon S., *Data mining for Marketing, Sales and Customer Support*, John Wiley & Sons.

2. Han Jiawei & Kamber Micheline, Data Mining: Concept & Techniques, Morgan Kaufmann, Elsevier Science, India
3. Johnson, P.E., "What Kind of expert should a system be?" *The Journal of Medicine and Philosophy*, vol 8, pp 77-97, 1983.
4. Konar, Amit, *Artificial Intelligence and Soft Computing: Behavioral & Cognitive Modeling of the Human Brain*, CRC Press, Boca Raton, Florida, 2000
5. McCarthy J. "Programs with common sense" as cited in Patterson (1990)
6. Patterson, D.W., *Introduction to Artificial Intelligence and Expert Systems*, Prentice-Hall, Englewood Cliffs, NJ.
7. Pujari, Arun K (2001), Data Mining Techniques, Universities Press, India
8. Russel, S. and Norvig, P., *Artificial Intelligence*, McGraw-Hill, New York, 1996
9. Waterman Donald A. *A Guide to Expert Systems*, Addison Wesley Long man Inc, U.S.
10. <https://www.javatpoint.com/application-of-ai>, site accessed on November 30, 2021

UNIT 15: EMERGING TRENDS IN IT

Structure

- 15.1 Introduction
- 15.2 Objectives
- 15.3 Competitiveness of ICT
- 15.4 Convergence of IT and Consumer Electronics: Emerging Trends
- 15.5 E-Commerce: Essential Components and Future issues
- 15.6 Privacy in the context of e-Commerce
- 15.7 Blockchain Technology and Cryptocurrency
- 15.8 Machine Learning, Big Data and Data Analytics
- 15.9 Summary
- 15.10 Unit End Exercises
- 15.11 References and Suggested Further Reading

15.1 INTRODUCTION

Information Technology, as you have seen in Unit 1 means the collection, storage, processing, dissemination, and use of Information. For dissemination of information, communication is essential. Persons involved in above activities must have proper skills for the specific tasks to have Information System developed, used, and information dissemination. IT departments of organisations should be aware of the technology for proper selection from the choices available for the set goals. At user level, be it personal or business, entire world knows today the importance of information and communication, much so, also because of social media and internet applications available through smart devices. Therefore tools and techniques are always sought to address the issues raised while attempting to utilize these two i.e. information and communication, for developing leverage.

Converged and unified communication has enabled faster communication. Convergence of data and voice network enables delivery of revenue generating services with more and more profits to the organizational business. The same message is delivered with multi channels viz. eMailing, SMS, Telephony, Social Media in unified way and hence brings higher satisfaction to the organizations and the customer both. Advent of digital switching systems and updates coming in terms of various versions in the components like hardware, software, communication media have been playing major roles in communication and information dissemination. As a result we find today much faster tools available to users at lower costs as compared to yester years.

Whenever the issue of convergence is discussed in the context of information technology it is essential to refer the convergence of information and communication systems so that devices can be unified for different application services. It is evident that the convergence of technologies will continue to be an important trend and the key to growth for various

sectors within the broader technology arena. The potential of new emerging technologies will play a key role in the development of new markets for new systems and with new and innovative ideas in automating processes.

E-commerce is the most widely acclaimed practice today. Almost every organization uses today its own website and portals for making business. The entire world has felt how IT was essential and helpful in the pandemic period for almost each and every services towards household commodities, online schooling, virtual meetings and so on. This was possible with active support of IT tools especially Internet.

This unit is an attempt to touch upon some of these aspects.

15.2 OBJECTIVES

After reading this unit you should be able to

- Describe the competitiveness of information and communication technology (ICT) with respect to modern business practices and their impact on economies;
- Describe technology issues involved in convergence with respect to software and hardware requirements and trends;
- Analyze global market penetration of ICT and prepare strategies for survival;
- Analyze emerging scenario in e-commerce; and
- Describe Application of AI, Data Analysis, and Big Data from business perspectives.

15.3 COMPETITIVENESS OF ICT

User of Information and communication technologies have proved as the tools for increasing efficiency levels in all spheres of operations. No organization can stand today without using ICT tools. For evaluating the impact of ICTs one has to go through the era known for high industrial growth i.e. latter half of the 20th century in which they have come into widespread use and have been going further with newer versions. The dramatic and continuing liberalization of world trade is indeed a central feature in global economy and the spread of ICT and goods derived using ICTs has been rapid with an accompanying emphasis on the issue of “competitiveness”. The word competitiveness is widely used to reflect the growing needs of business enterprises to evaluate their relative position with regard to their domestic competitors in international market and foreign competitors in domestic markets.

An enterprise is said be competitive if it can cut its share in international and domestic markets from its rivals. Similarly, the term competitiveness can also be used in the context of national economy to judge whether the nation is moving ahead in building market shares in business sectors that promote economic growth, employment, and other measures of social developments. The use or production of ICT is not itself, a principal determinant of social welfare measures like growth or employment, but it does play an important role in the ability of nations to participate in economic activities where

growth and employment prospects are more favorable than would be available otherwise. The use of ICTs strengthens the efforts of enterprises to achieve higher levels of competitiveness to contribute towards expansion of economic activities promoting growth, employment, and other social developments.

Historically, every major change in global economic activities has witnessed a complete transformation of industrial growth patterns involving different processes, practices and materials yielding explicit gains. In industrial nations, most of these gains had been realized by the first half of the 20th century owing to rapid industrialization. During the second half of the 20th century, the industrial growth of developing nations may have been constrained by the existence of the productive capacity of already industrialized nations. By the first half of the past century, most of the possible gains from this transformation had been realized in industrialized nations while developing nations, for many reasons, have experienced difficulties in fully entering this process. The ICTs provide an opportunity for addressing manufacturing productivity advances in developing nations, which would allow a substantial expansion of industrial output. Such output expansion, if large enough, could substantially influence the creation of employment and the creation of wealth.

Because ICTs are often laborsaving, a large increase in output is necessary to raise the derived demand for labor, through productivity improvement. Otherwise the impact of ICTs may be to reduce labor inputs, an undesirable and alarming outcome for developing nations if there are not sufficient employment alternatives. Much of this progress has been the direct result of the application of materials science to the production of modern ICTs. The earliest, and still the most important of these applications, is the use of semiconductors for the production of computers, telecommunications equipment, and related electronic devices. Technological improvement in semiconductor technology, as measured by the cost and speed of performing narrowly defined functions, has advanced at rates that were unprecedented for other manufacturing technologies such as power generation using fossil fuels or improvements in machine operating speeds through improvement in mechanical technology. The rapid advances in technical characteristics, which have also led to dramatic falls in price per unit of performance, have generated optimistic expectations about the contributions of information technology to manufacturing productivity growth.

Productivity gains from the use of information technologies involve improved control, smoother integration of production processes, and better control in the acquisition of inputs and the disposition of outputs. Communication technologies play important roles, particularly in co-ordination of functions, such as conveying timely information about inventories and scheduling throughout a distribution system. The productivity gains from ICTs are greatly realized through cost-reduction, in material inputs, labor, and capital. The relative shares of these reductions may differ across industries and over time, but labor saving is a principle source of cost-reduction. In addition, ICTs usually make it possible to produce more or higher quality with the same levels of inputs, resulting in productivity advances that are independent of changes in input use.

Growth in the use of ICTs in manufacturing appear to involve greater flexibility and changeover speeds as well as shortened and accelerated flows of materials for processing, work in process, and finished good inventories. These changes suggest a transformation in methods of organizing production systems from traditional models of mass production. The trade-off between the creation of new organizational models and the augmenting of old is essential for evaluating the issue of competitiveness. This is important for all the industries and sectors like Manufacturing, Education, Health, Pharmaceuticals, Hospital, Hospitality, Governments and so on.

The use of ICTs, have also made it possible to expand other economic activities, conventionally referred to as the “service sector”. The contributions of ICTs in the service area are apparent with the fact that ICTs are a major component in the predominant form of physical capital that service industries employ. It can be said without reservations that ICTs are the tools for productivity improvement in the post-industrial or information society age for achieving higher productivity gains that match or exceed those that have been historically experienced in manufacturing technologies. With such productivity gains it would be possible to indefinitely sustain the growth of economic output and productivity despite the trend towards a growing share of this output being produced in service sectors. The possibility of indefinite growth in the production and consumption of services is encouraging sign for both developed and developing economies.

For developed nations, an increasing share of output in services is consistent with increasing investments in human capital or the dispersal of industrial activities domestically and internationally to reduce localized environmental problems. The growth of services provide domestic growth opportunities that can absorb labor displaced by productivity improvements in agriculture and industry and that are less challenged by imports from developed or other developing nations. The corresponding problem for developing countries is to find ways to upgrade the value and quality of services so that service sector employees experience increasing wages over time. Doing this, of course, requires improvements in the productivity of the service sector. Thus, both developed and developing nations face a common challenge in finding ways to improve service sector productivity. Again, however, it is the issue of organizational change that provides a fundamental barrier to translating the rapid technological advances of ICTs into productivity gains. Developing and implementing the organizational changes would permit ICT use to have the same productivity impact in services as previous organizational changes had in manufacturing sector. For example, in financial services, where the ICTs do support economies of scale in transactions processing they also permit the creation of many new services.

There are several reasons to believe that developing nations could benefit from productivity improvements in services. First, in developing nations, governments often absorb relatively large shares of national output and most government activities are involved in the delivery of services. Hence, improvements in productivity in government services could free resources for private investment. Second, services are often close complements to manufacturing. For example, the effectiveness of the retail and distribution sector of an economy influences the growth of manufacturing by providing

more efficient market outlets for manufactured output. Third, developing nations are increasingly faced with the problems of harmonizing their production systems with the use of ICTs in developed nations so that they can serve as suppliers and sub-contractors in an increasingly global division of labor. This process of harmonization requires adoption of ICTs not only at the “service” level of the firm, such as the front office and the communication links to developed nation suppliers, but also within the production process to control quality and scheduling in ways that are consistent with customer demands. Many of these harmonization problems, nonetheless, are reflected in demands for services that, without the extensive use of ICTs, serve as barriers rather than complements to improvement in international trade and that absorb resources that could otherwise be used directly for production. Fourth, and finally, productivity improvements in both services and manufacturing are worthwhile wherever they may be achieved. Having more output using the same amount of inputs is of benefit in whichever sector it is achieved. To the extent that ICT use achieves greater productivity through releasing labor, the problem is to develop other opportunities for their employment (or remove barriers to this adjustment) rather than to lock them into employment patterns where they have low productivity.

ICT-intensive retailing activities, the systems that manufacturers employ must increasingly be compatible with the emerging new industrial models based on more intense ICT use. Examples of these developments include the growing emphasis on international quality standards with high levels of information content, the specification of product design using computer aided design and manufacturing, and the co-ordination of product delivery such as “Just in time” and “sales driven” production using high levels of ICT. Developing countries thus have an “offensive” strategic interest in adopting ICTs to maintain their competitiveness in export markets. Moreover, the liberal international trade environment that has characterized the “new world order” offers developing nations access to new markets, although developed countries maintain substantial import barriers in some areas, provided that they open their own markets to import competition. Imports from nations that employ ICTs to augment the flexibility and tighten control of the production process can offer formidable competition even with large differences in wage rates. Competition for domestic markets from exports in developing nations makes it necessary for developing country enterprises to adopt similar tools for achieving flexibility and variety that their competitors in developed and other developing nations are coming to employ. Effective use of ICTs requires knowledge from a range of disciplines and the solution of difficult problems of synthesizing technological knowledge and managerial “know how”. The absence of well-established models for effectively achieving this synthesis suggests that the creation of substantial variety of knowledge creating and using capabilities will improve the likelihood of adaptive success in realizing the potentials from ICT use. Many of the problems of implementing the new models of organization require skilled and flexible human resources that can solve a variety of problems.

15.4 CONVERGENCE OF IT AND CONSUMER ELECTRONICS: EMERGING TRENDS

Information is key to progress in decision making. It is so vital that it has created information age. The decision could be by a person or it could be automated by creating knowledge base from big data. Whenever the issue of convergence is discussed in the context of information technology it is imperative to refer the convergence of information and communication systems so that devices can be unified for different application services. The third dimension of human needs, which occupies the sphere of modern digital world with equal importance, is entertainment. The landmarks in history of the developments in communication are telephone (1870s), radio (1890s) and television (1930s) with increasing value addition for entertainment over information and communication. The personal computer (1980s) evolved primarily to address the information processing needs and it was the only device in the chain of revolutionary products developed which was digital by birth. The fast development in digital technologies paved the way of merger or hybridization of these different generation devices promoting the preferences for single device to cater information processing, communication and entertainment needs. This resulted in need driven efforts for convergence of technologies with an eye to develop innovative consumer products which can provide a common platform for handling different services and applications. The mix of research & development of technologies and innovative product development and design created a new paradigm to ever-increasing urge for value-added products. This process of value addition is creating new dimensions to existing products and setting newer trends for consumer preferences and needs well supported by convergence of technologies.

Seeing the trend in the growth of electronic components used in computers, Gordon E. Moore gave a prediction, way back in 1965 that the number of transistors per silicon chip doubles each year. This is known as Moore's Law. In all the computers and smart devices silicon chips are used. Innovations and further developments have resulted in bringing down the size of the actual product. The concept of Large Scale Integration (LSI) and further Very Large Scale Integration (VLSI) came into existence because of which we find today small computing devices and IT tools with tremendous capabilities.

The world has seen the convergence of technologies. It will continue to be an important trend and the key to growth for various sectors within the broader technology arena. Use of internet has penetrated in all the sectors. Automation with Internet of Things (IoT) can be seen with tremendous applications. The potential of new emerging technologies will further play a key role in the development of new markets for new systems and with new and innovative consumer electronic products. The innovations are more rapidly seen in home networking, home appliances, games, Blue tooth devices, digital imaging, digital entertainment platforms, wireless technologies, and digital radio receivers. Similar is the case with gaming, defense, agriculture, transport, and hospital industries.

One major aspect related to the above innovative technologies is to develop seamless wireless connections between small, multi-functional devices in the home. The signal-processing technology is playing tremendous role for digital audio, video, cell phones, machine learning, broadband and wireless communications. In fact, signal processing has given a number of tools for deep learning and statistical analysis. Data Science has become a separate discipline for mining, analyzing, and decision making. The expansion of broadband and high-speed wireless networks have proved to be a key factor in the present and seems to continue in future proliferation of IT products to be used by common people in daily life. Fiber connectivity at home with internet applications like iTV clubbed with blue tooth devices are providing endless services. It is expected that more content made available over such networks and more engaging applications for the user will lead to the development of larger commercial markets. Most of the advances in this direction will lead to developments in consumer electronics products in which convergence of different technologies and applications will come up as natural process.

The consumer electronics industry, in short, is entering a new era based on the promise of utilizing broadband and wireless technologies to promote the more widespread use of electronic devices. Advanced in microelectronics further enhances the potential for consumer electronic devices to proliferate and be utilized well beyond what is in the market today. The personal computer led to a revolution in terms of how data could be managed and manipulated. The Internet enhanced this function by providing direct access to more data. Mobile wireless and wideband provide the consumer with all of the above in more places with greater access. The trend is leading many markets that were traditionally separate on a competitive collision course.

Convergence trend has also been leading both hardware and software companies to become both hardware and software companies. Systems are open to share contents and associated resources. Digital technology is going further with developments keeping in view the ease of operations with more and more features made available to users. Smart phones and devices with Android and iOS are playing major roles in connecting person with friends, family, and groups globally. This trend will further go with more features and with the device size. We find desktop computers, laptops, and all other hand-held devices such as tablets and smart-phones, useful in transmitting downloaded content to the TV. In fact we also find TV bypassing the computers entirely and getting its content by using set-top box supplied by the internet service providers. Video on demand is possible at anytime with just at a click of button. Such a trend of convergence has provided an advantage to hardware makers over software companies and those that excel at branding and the farming out production to other firms.

New emerging technologies feature new, innovative products with a focus on the convergence and consumer adoption of 4G, highlighting the growth potential after the success of 3G and GSM. 5G is already available in developed countries and going to be adopted by other countries also in near future. India has already announced for it's implementation in phased manner. In such countries wireless phone owners prefer converged wireless phone/PDA devices to multiple stand-alone products capable of

performing the same functions. More advanced phones are made for more advanced networks. Many advances in wireless technologies are seen for next-generation wireless systems such as smart phones with high-definition screens, multiple camera functionalities and those with slim sizes and secured features and facilities. This has resulted due to better, faster networks being setup and strong demand from consumers.

In the field of communication, the public switched telephone network (PSTN) was designed to accommodate a single application i.e. the voice communication. It was not built for services such as data and video. During the late 1980s and early 1990s, the rapid growth of distributed processing and the Internet changed the telephony world. Internet congestion overloaded PSTN circuits, and carriers (service providers) were forced to invest in additional PSTN capacity which was expensive and cumbersome task. The characteristics and bandwidth demands of data transmission were not suited for the circuit-switched PSTN. The packet-switched network accommodated the rapid growth and bandwidth requirements of data traffic more efficiently than the circuit-switched architecture. As carriers developed the packet-based infrastructure to support the increasingly data-centric traffic, the next logical step was to carry voice traffic on that same network. The convergence of voice and data networks enabled the delivery of advanced, revenue generating services, which were not feasible in the circuit-switched environment. TCP/IP protocols with advancements in addressing techniques have played significant role in internet applications. It will further require innovative developments to take care of secured communications and transactions over internet.

Network signaling is critical to the delivery of services in converged networks. The standard signaling technology, Signaling System 7 (SS7), allows PSTN network nodes to communicate directly with one another and to share the information needed to process any type of call autonomously. SS7 is a reliable, self-healing network architecture and protocol that manages voice circuit functions on a separate, fully redundant packet-based network. It is widely deployed in both wire line and wireless networks worldwide. In addition to managing call set up and tear down, SS7 enables services such as toll free, and toll wire line services, and local number portability (LNP). Enhanced call features - like call forwarding, calling name delivery, and three-way calling, as well as business systems, including customer care and billing, are all enabled by SS7 signaling. Likewise, services such as Internet call waiting, prepaid calling, calling party pays, Internet roaming, and wireless web browsing requires the support of SS7 signaling. Users have the ability today to control how, when, and with whom they communicate. The new generation of services will focus on secured communication with personal identity, mobility, presence management, and multimedia/multiparty communication with virtual reality. The ability to mine the signaling data to support business systems such as billing and fraud control will prevail as a requirement.

The generations in wireless spectrum from GSM to 3G to 4G and now 5G have provided communication through airwaves. Each generation provides better speed to share data, voice, and videos over the network. The latest

being 5G is supposed to be nearly 100 times faster than 4G with super-high-frequency airwaves.

Virtual Reality (VR) and Augmented Reality (AR) are another applications which will have more penetration in the fields like Gaming, Education, and Businesses. Augmented Reality simulates artificial objects in the real environment whereas Virtual Reality creates an artificial environment to stay in the place using simulation. The technologies will go a long way to make a profound impact.

15.5 E-COMMERCE: ESSENTIAL COMPONENTS AND FUTURE ISSUES

E-commerce is widely used application of internet for business organizations. It is the most acclaimed buzzword, which gained popularity even aftermath of so-called 'dot com' technology boom and diffusion in adopting web technology. Every business today necessarily requires active support of IT tools especially internet for maintaining web sites and for the unified communications. Though various business models evolved and still the process of finding the most suitable model with secured access for different business propositions is continuing, the impact of e-commerce practices can be felt and acknowledged without any reservations. Use of this has been experienced in the pandemic period for online services towards purchase of consumer goods, household goods, entertainment, gaming, education and many more for survival.

This section provides a brief description of modern practices and emerging trends related to technology, design and security issues involved in e-Commerce.

Wireless Internet

All business companies in the field of communication are developing and marketing wireless communications services and products required for facilitating business through wireless internet. Instant unified messaging are available to customers and others involved for connecting faster. Anything you can do on the Internet can be accessed via your WAP(wireless application protocol) cell phone. WML (Wireless Mark-up Language) has been designed for devices with small displays. It is based on XML (eXtensible Mark-up Language). The focus now is on the development of open, generic software products and platforms based on the WAP standard.

Bluetooth

Bluetooth is a wireless technology which is used low-cost, short range radio links between mobile PCs, mobile phones, PDAs, and other portable devices. By using Bluetooth, users can connect a wider range of telecommunication and computing devices without cable giving much relief to organizations and users. Data updated on one device is automatically transferred to all other connected devices seamlessly.

Portals

Portals are the organisation's web sites for making business. They combine a portfolio of basic content, communication, and commerce sites. For the

most part, they started out as search engines. There are mainly two types of portals – Vertical, and Horizontal portals. However, With the wide range of applications, however, they can be categorized in different types of portals. Vertical portals are the sites that focus on a particular content category, commerce opportunity, or audience segment, with a broad set of services. Examples of such portals are Chem Industry, Construction Plus, eBay etc. Horizontal portals are web portals which focus on wide range of topics and interests for general audience for example google.com, msn.com etc. Marketplace portals support business-to-business and business-to-customers e-Commerce. They also provide software support for e-Commerce transactions with ability to find and access information about the products on sale. Further categories could be Geographical portals (countryweb.com, craigslist.com), Access portals (aol.com, orange.co.uk), Media portals (bbc.co.uk, guardian.co.uk), Search portals (ask.com, google.com, bing.com), Payment portals (razorpay.com, PayPal, PayTm, GooglePay) etc.

As the website of the business organization offering e commerce services are their front offices and showcases for their customers, a good design can make a web site stand out among dozens of competing sites selling similar products. A well-thought-out presentation can boost traffic and sales. The best-designed sites keep customers coming back. Simplicity seems to be the best strategy in web design for e-commerce. E Commerce managers and web-design staff must ensure that it's easy for users to download a page, find an item, make a purchase, and navigate through page after page without losing track of the starting point. Navigation is seen as the key; the first step to building a simple yet effective e-commerce site.

On Line shopping

The biggest hurdle in online shopping faced by the customers is lack of opportunity to physically verify the suitability of the product for his or her needs. In some category of products it becomes almost impossible to make a decision based on the passive display of product. Innovations in these directions are aimed to remove this hurdle and some of the sites have shown a path to resolve this problem. IT tools for direct response applications should be used extensively to allow merchants and retailers to offer their customers an interactive shopping experience based upon scenes, photos or pictures that contain embedded, contextual advertising and merchandising links. The approaches that allow customers to see the products in real time before they purchase are required to be encouraged.

Catering websites to customers' preferences increases the depth of knowledge about customers, reduces customer-acquisition costs, and builds brand loyalty. This is known as site personalization. Personalization means presenting the site visitor with specific information based on his or her identity or behavior. It is all about making the content more relevant for the user. The tools for personalization include click stream tools, which watches the pages that the visitor browses and keeps track of time spent. Thus they know where the customer has been on the site. Others tools such as collaborative filtering tools examine data based on explicit preferences, and statistics about likely purchases come from the patterns shared with similar users in the database.

Rules-based systems build user profiles and incorporate business rules driven by the site owner (i.e. rule can instruct server to show the site in French if the visitor comes from a French domain) and also barriers which include price and privacy concerns.

Live customer service companies are responding to customers' requests for customer service throughout 24 hours 7 days a week. One of the ways by which they have done this is to create a virtual assistant to ask questions like a real person would ask, and as a result, gets/delivers email or instant live messages from/to the customers. In case customers want to talk to a real person, the request is executed accordingly. E-businesses are realizing that customer service is the bottom line when it comes to how successful they will be.

15.6 PRIVACY IN THE CONTEXT OF E-COMMERCE

One of the most controversial topics in e-commerce today is the privacy of customers. The most common apprehensions are like, what are they doing with our information? Why do they want it? What will they do to our so-called private lives? In order for businesses to build lasting and fulfilling relationships with customers, they need to know as much about them as possible. However, as all of us are aware, this information is being used for purposes other than to serve us. The big question for e-businesses is how to collect their customers' information yet still maintain their trust. This has been an issue right from the beginning and will go further in future. E-businesses will have to learn to balance personalization with privacy issues. Companies must only gather and use data that will help their customers complete the business on their website. However, the reality is that information is routinely being sold and used for purposes other than the reason the information was originally gathered. So what are the organizations' responsibilities?

Personalization on the Web takes two basic forms: collaborative filtering and profiling. Both are the attempts to collect specific information that will be valuable to generating additional traffic, click-throughs, and revenue on Web sites. Collaborative filtering is voluntary. It takes information provided by the user and predicts what information will be of interest to him or her. It can also compare a user profile with those of other users to create common-interest groups and make recommendations based on other customer preferences. In profiling, a visitor's actions are observed and gathered through the use of a cookie, a unique identifier for that customer. Subsequent user clicks identify patterns of behavior and interests, which result in the presentation of specific information to Web sites visited. There is also the ability to allow the collection of information on an individual's cookie across multiple sites.

Open source model for customer data

CPEX (Customer Profile Exchange) is a standard which is followed by e-business companies to gather customer information for further used by them for their respective business. The standard has been developed by a group of large companies making it to be followed for e-business. It

facilitates the privacy-enabled exchange of customer profile information. The CPExchange specification defines a data model for profile information that relates a customer or other business partner, to an enterprise. The CPExchange specification also defines metadata for associating privacy controls with various subsets of the profile information, as well as operations for query, delivery, and update of this information. This information is represented in Extensible Markup Language (XML), which can be exchanged through messages.

Customer service issues

Bad customer service has a negative effect on e Businesses. Technical difficulties with shopping carts, failure of customer service representatives to respond to inquiries and frustrations about returning items are among the issues that customers often complain about.

Standard for e-commerce

A need for a standard in e-commerce has been recognized and the standard for Internet commerce was developed by Global Information Infrastructure (GII) in association with Ziff Davis. GII is defined as a seamless web of interactive communications developed at world-wide level to provide the infrastructure for new services and activities based on the strategic use of all types of information. The Commerce Standard is a codification of best practices in Internet commerce developed by a community of world-leading thinkers and practitioners; it specifies the merchant practices and policies that lead to high levels of customer satisfaction, service, security and privacy. The Standard for Internet Commerce meets five primary needs. Together these five needs make it crucial that the Commerce Standard is developed and implemented. The Standard for Internet Commerce is needed:

1. To increase consumer satisfaction and confidence in doing business on the Internet by merchants and customers as a way to establish merchant credibility and trustworthiness,
2. To help merchants provide a world-class customer experience, innovate rapidly and lower their costs,
3. To support and enhance self-regulation of Internet commerce, and
4. To help merchants and customers deal with a proliferation of guidelines and symbols.
5. To help merchants and customers follow the norms set up by the authorities and regulatory bodies to safeguard the business transactions and customer interests.

15.7 BLOCKCHAIN TECHNOLOGY AND CRYPTOCURRENCY

A typical database stores data on a particular subject accessible to users. The users of this database could be at one place in the organization or scattered over different regions. Distributed databases are those available at different places and accessible through networks for sharing the content by authorized users. Blockchain technology is based on the concept of distributed database. It is a structure that stores transactional records of

the public in several databases connected over the network of users. The transactional records are called Blocks. The databases are known as the “chain,” in the network connected through peer-to-peer nodes. Typically, this storage is referred to as a ‘digital ledger’.

Every transaction in this ledger is authorized by the digital signature of the owner, which authenticates the transaction and safeguards it from tampering. This guarantees the safety of the information contained in the digital ledger. To understand better, the digital ledger is like a Google spreadsheet shared among numerous computers in a network, in which, the transactional records are stored based on the actual activities performed. The activities could be related to actual purchases made, items issued from stores on a day and so on. The interesting part is that anybody in the group can see the data, but they can’t alter or corrupt it.

To be more precise, a blockchain is shared among the nodes of a computer network. As a database, a blockchain stores information electronically in digital format. Blockchains are best known for maintaining a secure and decentralized record of transactions using digital currencies. Blockchain is based on trust without the need for a trusted third party.

One key difference between a typical database and a blockchain is how the data is structured. A blockchain collects information together in groups, known as blocks, that hold sets of encrypted information related to the transactions. This information is mainly about digital key, date and time and other related information. Each block has certain storage capacities. When the transaction is complete then block is closed and linked to the previously filled block. This forms a chain of data known as the blockchain. The same steps are performed by one and all in the group.

A typical database usually structures its data into tables. A blockchain structures its data into chunks (blocks) that are strung together. This data structure inherently makes an irreversible time line of data when implemented in a decentralized nature. Each block in the chain is given an exact time stamp when it is added to the chain.

The goal of blockchain is to allow digital information to be recorded and distributed, but not edited. In this way, a blockchain is the foundation for immutable ledgers, or records of transactions that cannot be altered, deleted, or destroyed. This is why blockchains are also known as a Distributed Ledger Technology (DLT).

Because of the decentralized nature of blockchain, all transactions can be transparently viewed by either having a personal node or using ‘blockchain explorers’ that allow anyone to see transactions occurring live. Each personal node has its own copy of the chain. Chain gets updated as new authenticated blocks are confirmed and added. Since records stored in the blockchain are encrypted, only the owner of a record can decrypt it to reveal their identity. This is done by using a public-private key pair.

This industry has come up by investing millions by leading companies in convening networks, establishing governance models, developing flexible blockchain technology and more. It is easy for those who want to join others in the ongoing movement to transform the food supply, supply chains, trade finance, cross-border payments and many others.

IBM Blockchain Services is the industry's leading blockchain services provider with more than 500 clients projects, backed by thousands of technical and business experts.

Processing steps of blockchain are illustrated in the following Figure 15.1.

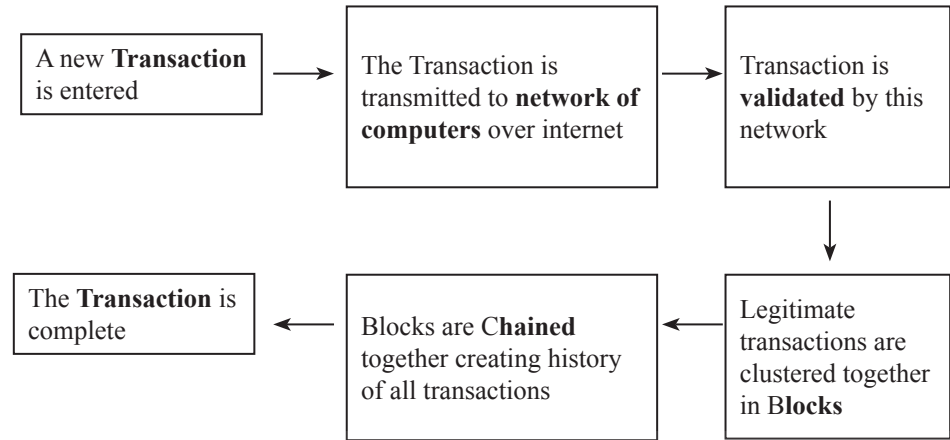


Figure 15.1 : Blockchain Processing Steps

The currency used in blockchain is named as **Cryptocurrency**. It is digital currency that can exist in any form, digitally or virtually. It uses cryptography to secure transactions from peer-to-peer without involving any bank or regulatory mechanism. It is not in terms of physical money to be carried out for exchange. They are purely as digital entries to an online database. When you transfer in cryptocurrency then the transactions are recorded in a public ledger. Cryptocurrency is stored in digital wallets. Because of cryptography used to verify transactions, it has been named as cryptocurrency. Advanced coding is involved in storing and making transactions to provide security and safety. Units of cryptocurrency are created through a process called mining. This process involves computers to solve complicated mathematical problems by using encryptions to finally generate digital coins. Users can also buy the currencies from brokers, then store and spend those using cryptographic wallets. If you own cryptocurrency, you don't own anything tangible. What you own is a key that allows you to move a record or a unit of measure from one person to another in the group without any third party.

The first cryptocurrency was **Bitcoin**, which was founded in 2009 and remains the best known today. Other main cryptocurrencies, as of now, are Ethereum (founded in 2015), Litecoin (founded in 2011), and Ripple (founded in 2012). Ripple can be used to track different kinds of transactions, not just cryptocurrency. Non-Bitcoin cryptocurrencies are collectively known as "altcoins" to distinguish them from the original.

Bitcoin uses, as stated above, Peer-To-Peer technology to operate with no central authority or banks. It is open source and can be accessed by all. Nobody owns or controls Bitcoin. Managing transactions and issuing of Bitcoin is carried out collectively by the network. Transactions are secured by cryptographic signatures preventing other people from spending your money. Proof-of-Work (PoW) accomplished by you prevents other people from undoing, rearranging or losing your transactions. Sending bitcoins across the borders is as easy as sending them next door without waiting time.

and without paying any extra fee towards transactions for any international transfer. There is no limit on minimum or maximum amount.

15.8 MACHINE LEARNING, BIG DATA AND DATA ANALYTICS: FUTURE TRENDS

In previous Unit, you have learnt about role being played by Artificial Intelligence in decision making. In AI perspectives, buzz words these days, one comes across, are machine learning, deep learning, big data, data analytics. We will discuss how these are related to each other and how they are being put in use by organizations in automating processes thus giving further opportunities.

Learning is defined as getting knowledge. In other words, knowledge is created by learning. It is the process of developing skill. Machine learning is termed as creating knowledge in machine. The machine could be a computer, smart device, or simply a device attached to an object for automating some processes. Machine learning is a branch of Artificial Intelligence. It deals with effectively automating the process of doing data analysis and allowing the machines to adapt to new scenarios independently. Analyzing huge amounts of data available through internet and providing critical insights to the business owners to provide them with a larger picture, with the help of which they can make informed decisions, is done through machine learning. With the application of AI, machines are made capable of making decisions for businesses autonomously to achieve the highest Return On Investment (ROI).

With neural network a human brain takes decision. In other words, consciousness is created by the neural network. Similarly, artificial neural networks are used in creating intelligence in machines for taking decisions. Let us see now how machine learning and deep learning are related.

Machine learning is the broader category of algorithms that are able to take data on a particular subject. The data could be from all possible sources available within or beyond the organization. The basic purpose of these algorithms is to identify patterns, discover insights, and/or make predictions based on the analysis. Deep learning is a particular branch of machine learning that takes ML's functionality and moves beyond its capabilities.

With machine learning in general, there is some human involvement to review the outcomes. On the other hand, deep learning doesn't rely on human involvement for review. Instead, a deep learning algorithm uses its own consciousness created by more sophisticated artificial neural network to check the accuracy of its results and then learn from them. The neural network in deep learning learns on its own as to how to get better at a task over time without any feedback given by human involvement.

There are many practical advantages of using machine learning. In healthcare we find image recognition and analysis for diagnostic purposes thereby saving lives. Robots are in use for doing surgery. Similarly CT scans, MRI (Magnetic Resonance Imaging) are in use for diagnosis and corrective treatments of life threatening diseases.

We will see some more advantages of ML for better understanding. They could be broadly categorized as big data management, creating intelligent

devices, and creating consumer experiences. They are briefly mentioned below. Based on applications being added, it is predicted that by the year 2030, around 70% of businesses will be using at least one type of AI technology and roughly half of all large companies will have a full range of AI tech embedded in their process.

Big data management – Big data is mined by using various sites on internet, apart from the data available in the organization. Mining, gathering, and doing analysis on big data is a massive task which is just not possible without the speed and sophistication of machine learning. Machine learning algorithms are devised and coded in order to perform these tasks. You can imagine the huge volume of data momentarily being generated over the globe with human interactions by using digital media.

Intelligent devices – Use of IoT (Internet of Things) is facilitating in developing wearable devices to track health and fitness parameters and goals set by the individuals. In automobiles you find self-driven cars which can reduce wasted time and energy. Similar applications are that of developing drones to deliver goods, generate patterns by clicking automated cameras and so on. In household goods too you find smart appliances performing some or the other task on their own without involving human. IoT holds great promise, and machine learning can help make sense of this significant increase in data.

Creating consumer experiences – Every established business organization uses its websites and portals for making businesses. All these sites have search engines for providing assistance to their customers or interested persons in searching items or any other information. The web browsers have strong search engines for providing data from various sites based on the keywords entered by the users. This is also available by enabling sound devices and providing results using the voice inputs. Technical tools like “Assistants” are available to listen and respond to the voice input provided. Machine learning enables search engines, web apps, and other technology to customize results and recommendations to match user preferences, creating personalized experiences for consumers.

Figure 15.2 gives an overall view of the ML process starting from data input from the data resource to create the information base for further processing to have knowledge base. Creating wisdom in machines for supporting the decision making process in organization is the ultimate product.

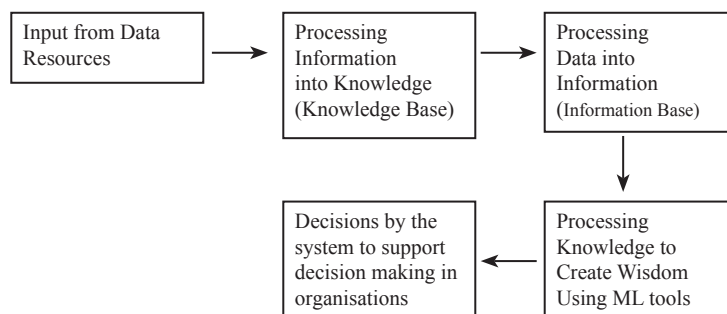


Figure 15.2: Decision support using Machine Learning

To be precise more about use of ML in above categories we find today banks using ML predictive models to better understand and meet customer

needs, to get analysed reports minimizing the risk in business, to identify cyber threats, track and document fraudulent customer behavior, and better predict risk for new products. ML in banking services also includes fraud detection and mitigation, personal financial advisor services, and credit scoring and loan analysis.

Similarly, manufacturing companies use automation to optimize production by utilizing resources in better and improved way of scheduling the related resources. Root cause analysis, supply chain and inventory management are further applications in manufacturing which ML modeling is being used for.

Let us take some examples of the ML modeling systems. ‘Infervision’ company has developed an AI and trained with suitable algorithms to review the CT scans and detect any early signs of lung cancer. It makes the job easier for radiologists because they can simply use the data from the AI and can diagnose for cancer more accurately and efficiently.

‘MakerSights’ use a product decision engine for retail sector of making businesses. It supports informed decision-making throughout the product creation, work-in-progress and go-to-market process.

The automobile industry has been making good use of IoT and AI by using increasing number of sensors for security purposes and for recording the vehicle movements.

Volvo uses IoT and AI to uphold and maintain its safety reputation. In a project launched in 2015, Volvo fitted 1,000 cars with sensors to detect and analyze driving conditions and to monitor the vehicle’s performance in hazardous conditions. Vehicles fitted with automated features generate good amount of data. The collected data is then uploaded to their cloud. Volvo works on this data for carrying out machine-learning driven analysis across its collected data. Volvo’s early warning system analyzes over a million events per week to predict breakdown and failure in their cars.

British Petroleum operates in 72 countries worldwide. The giant in oil and gas has been leading the industry with the benefits of big data and AI in its processes. Sensors installed in its gas and oil wells continuously collect data to monitor and understand the working conditions of the well at each site irrespective of the physical location. Analyzing this data helps in monitoring and optimizing the equipments performance and their maintenance needs to enable smooth and unhindered functioning.

‘Underwrite.ai’ is used to access credit risk for small businesses and consumer loan applications by applying advanced artificial intelligence algorithms to acquire portfolio data to find patterns for good and bad loan applications.

Using AI in business has given the term **Business Intelligence (BI)**. Over the past decade, business intelligence has been revolutionized. Data exploded and became big. And just like that, we all gained access to the cloud. BI use is evolving. Business houses are interested about ‘what is the best BI solution for their specific business’. Dashboards are helpful to see what is happening at every second. They give alerts in case something is undesirable and unexpected. Data discovery with clean and secure data combined with a simple and powerful presentation will be in demand in near future.

Demand for real-time online data analysis tools is increasing. With IoT (Internet of Things) an uncountable amount of data is being generated. Data so generated requires statistical analysis and management at the top of the priorities list. Embedded analytics is becoming a standard in business operations. Departments and company owners are looking for professional solutions to present their data without the need to build their own software.

Data Analytics (DA) is the process of examining data sets in order to find trends and draw conclusions about the information they contain. It is the science of analyzing raw data to make conclusions. The information obtained can then be used to optimize processes to increase the overall efficiency of a business or system.

For example, manufacturing companies often record the runtime, downtime, and work queue for various machines and then analyze the data so obtained to better plan the workloads helping the companies to put the machines in operation to its maximum capacity.

Unlike business analytics where data analysis is oriented to business use, data analytics refers to variety of applications, from BI, reporting and online analytical processing (OLAP) to various forms of advanced analytics. Hence it has a broader focus.

Data analytics initiatives can help business companies in increasing revenue, improving operational efficiency, optimizing marketing efforts apart from supports for customer services. Analytics also enable organizations to respond quickly to emerging market trends and gain a competitive edge over competitors and business rivals. The ultimate goal of data analytics, however, is boosting business performance.

Data analytics can be separated into quantitative data analysis and qualitative data analysis. The former involves the analysis of numerical data with set of variables. These variables can be compared or measured statistically. The qualitative approach focuses on understanding the content of non-numerical data like text, images, audio and video, as well as common phrases, themes and points of view.

When data analytics occurs within a user's natural workflow, it is called embedded analytics. Businesses have recognized the potential of embedding various BI components such as dashboards or reports into their own application and thus improving their decision-making processes and increasing productivity.

In Gaming industry data analytics is used to set reward schedules for players. This helps keeping the majority of players active in the game. Similarly, companies dealing with content may use data analytics to keep you clicking, watching, or re-organizing content to get another view or another click.

Data analytics is a broad field. There are four primary types of data analytics:

- Descriptive analytics,
- Diagnostic analytics,
- Predictive analytics, and
- Prescriptive analytics.

Descriptive analytics techniques help in answering questions about what happened over a given period of time. Other questions could be ‘Have the number of views gone up? Are sales more this month than the last month?’ and so on. The techniques summarize large datasets to describe outcomes to stakeholders. The process requires the collection of relevant data, processing of the data, analyzing, and presenting to users.

Diagnostic analytics focuses more on why something happened. Did the weather affect seasonal sales? Did that latest marketing effort impact sales? These techniques take the findings from descriptive analytics and dig deeper to find the cause. The parameters affecting performance are investigated to discover why the performance got better or worse. To achieve the outcomes generally occurs in three steps are involved – to identify anomalies in the data, collecting the data related to these anomalies, and applying statistical techniques to find relationships and trends that explain these anomalies.

Predictive analytics focuses on what is likely to happen in the future. What happened to sales when we had hot summer last time? How many weather models predict cold wave this year? The techniques use historical data to identify trends and determine if they are likely to reoccur. This is achieved by using a variety of statistical and machine learning techniques, such as: neural networks, decision trees, and regression.

Prescriptive analytics focuses on what should be done, by suggesting a course of action. Prescriptive analytics techniques rely on machine learning strategies that can find patterns in large datasets. By analyzing past decisions and events, the likelihood of different outcomes can be estimated.

Prescriptive analytics goes a step further into the future. It examines data or content to determine what decisions should be made and which steps are taken to achieve an intended goal. Prescriptive analytics can help you optimize scheduling, production, inventory, and supply chain design to deliver what your customers want in the most optimized way. These are some of the emerging trends in business intelligence.

To make you aware more about the data analytics, two examples are given here. They belong to predictive analytics category. These two analytics methods are quite popular among data scientists: Artificial Neural Networks (ANN) and Auto Regressive Integrated Moving Average (ARIMA).

In Artificial Neural Networks, data is being processed in a similar way as in biological neurons. In the human brain, power of neural networks lies in their capability to connect sets of neurons together in layers and create a multidimensional network. The input to the second layer is from the output of the first layer, and the situation repeats itself with every layer. This procedure allows for capturing associations or discovering regularities within a set of patterns with the considerable volume, number of variables, or diversity of the data.

ARIMA is used for time series analysis. It takes data from the past to model the existing data and make predictions about the future. The analysis includes inspection of the correlations by comparing current data values and past data values. How are the current data values depend on past values so that they could be considered in making predictions. ARIMA techniques are complex but provide a very powerful tool for predictive analysis.

Data analytics initiatives support a wide variety of business uses. For example, banks and credit card companies analyze withdrawal and spending patterns to prevent fraud and identity theft. Transactional logs and click-stream are used by companies in e-Commerce to identify website visitors who are likely to buy a particular product or service, based on navigation and page-viewing patterns. Healthcare organizations use patient data to evaluate the effectiveness of treatments for cancer and other diseases.

The analytics process starts with data collection. The next step is integrating and preparing data for developing the analytical model. The last step is testing and revising the analytical model to ensure that it produces accurate results. In addition to data scientists and data analysts, the analytics teams often include data engineers, who create data pipelines and help prepare data sets for analysis. Data collected from different sources may need to be combined via data integration routines. The integrated data is transformed into a common format and loaded into an analytics system, such as a Hadoop cluster, NoSQL database or data warehouse. Programming languages such as Python, Scala, R and SQL are used in the process. The last step in the data analytics process is communicating the results generated by analytical models to business executives and other end users. Various charts and other info-graphics can be designed to make findings easier to understand. Figure 15.3 below illustrates the process of data collection, integration, data analytics, and presentation of end results.

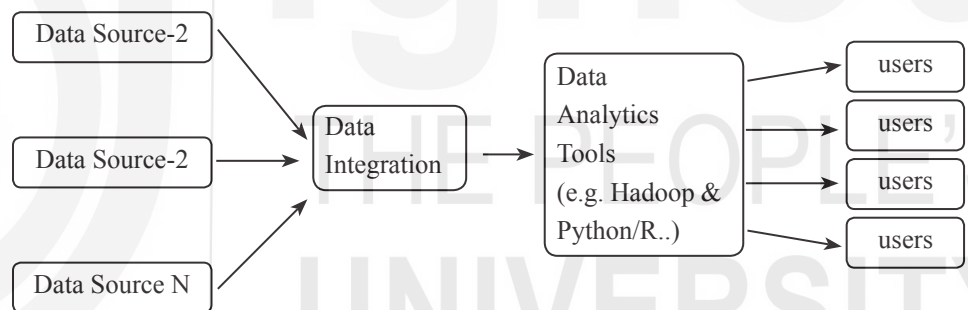


Figure 15.3 Data Collection, Integration, Analytics and Presentation Process

Data analyst's skills combine three fields – IT, Statistics, and Business in order to help businesses and organizations succeed. The primary goal of a data analyst is to increase efficiency and improve performance by discovering patterns in data. The task involves working with data throughout the data analysis pipeline.

Data analytics has an important role in the banking and finance industries. It is used to predict market trends and assess risk. Credit scores are an example of data analytics that affects everyone. These scores use many data points to determine lending risk. Data analytics is also used to detect and prevent fraud to improve efficiency and reduce risk for financial institutions. Crime preventions, environmental protection, health informatics are other areas where data analytics can provide critical information. Pharmaceutical companies also use data analytics to understand the market for drugs and predict their sales.

With the availability of market-leading decision-making software available, businesses are increasingly shifting towards the notion of

bringing planning and analytics together for complete visibility across their activities. Guiding this shift is Integrated Business Planning (IBP), an approach to planning that unifies financial and operational data from across the organization.

15.9 SUMMARY

Multimedia, wireless networks, IoT, machine learning, applying data analytics to businesses, and other convergences define the new focus of ICT. The converged network combines voice services with packet networks, using signaling as the underlying technology. Whenever the issue of convergence is discussed in the context of information technology it is essential to refer the convergence of information and communication systems so that devices can be unified for different application services. Emerging protocols and new network architectures are developing to support changing technologies and subscriber demand for new services. Major technology and business companies are in the lead in developing and marketing wireless communications services and products required for facilitating business through wireless internet. Machine learning and IoT applications are generating big data for business and data analytics for improving services in all the fields. The ICT tools provide an opportunity for addressing productivity advances in nations, which would allow a substantial expansion of industrial output.

This Unit has highlighted the competitiveness of information and communication technology (ICT) with respect to modern business practices. Big data generation, data analytics and machine learning have been introduced in the Unit. The Unit described the technology issues involved in convergence with respect to software and hardware requirements and trends. E-commerce is coming up in big way. The Unit described modern practices and emerging trends related to technology, design and security issues involved in e-commerce. Entire world has experienced the online facilities leveraging the life style in the pandemic period. Blockchain technology is coming up and being adopted by more and more countries. The Unit has also covered the emerging trends that has arisen due to convergence of IT and consumer electronics.

15.10 UNIT END EXERCISES

1. It is said, “Information and communication are most of the time complementary and sometimes supplementary to each other”. Elaborate!
2. What do you understand by competitiveness of ICT? Highlight the impact of ICT for the developing nations with respect to the economic activities.
3. Write a detailed note on emerging trends in convergence of IT and consumer electronics.
4. Briefly describe the modern practices and emerging trends related to technology, design and security issues involved in e-commerce.
5. What do you understand by Blockchain? Describe its advantages.

6. Write a short note on Cryptocurrency.
7. Write short notes on Machine Learning, and Data Analytics.

15.11 REFERENCES AND SUGGESTED FURTHER READING

Dorothy E. Leidner, Robert D. Galliers, D. E. Leidner, *Strategic Information Management: Challenges and Strategies in Managing Information Systems, Third Edition*, Butterworth-Heinemann.

John Ward, Joe Peppard, *Strategic Planning for Information Systems*, John Wiley & Sons; Third edition.

Steven Shepard, *Telecom Convergence: How to Profit from the Convergence of Technologies, Services, and Companies*, McGraw-Hill Professional Publishing; 2nd edition.

<https://www.computermachinelearn.com/> site accessed on February 23, 2022.

<https://www.hpe.com/in/en/what-is/machine-learning.html> site accessed on February 23, 2022.

<https://www.investopedia.com/terms/d/data-analytics.asp> site accessed on February 21, 2022.

<https://www.mastersindatascience.org/learning/what-is-data-analytics/> site accessed on February 21, 2022.

<https://searchdatamanagement.techtarget.com/definition/data-analytics> site accessed on February 21, 2022.

<https://academy.binance.com/en/start-here> site accessed on February 15, 2022.

<https://www.kaspersky.com/resource-center/definitions/what-is-cryptocurrency> site accessed on February 21, 2022.

<https://towardsdatascience.com/machine-learning-and-signal-processing-103281d27c4b> site accessed of February 13, 2022.

<https://linchpinseo.com/trends-in-the-information-technology-industry/> site accessed on February 12, 2022.

<https://www.managementstudyguide.com/emerging-trends-in-information-technology.htm> site accessed on February 12, 2022.